Fifth Quarterly Progress Report

October 1 through December 31, 1999

NIH Project N01-DC-8-2105

Speech Processors for Auditory Prostheses

Prepared by

Blake Wilson, Dewey Lawson, Robert Wolford and Stefan Brill

Center for Auditory Prosthesis Research
Research Triangle Institute
Research Triangle Park, NC 27709

I. Introduction

The main objective of this project is to design, develop, and evaluate speech processors for implantable auditory prostheses. Ideally, such processors will represent the information content of speech in a way that can be perceived and utilized by implant patients. An additional objective is to record responses of the auditory nerve to a variety of electrical stimuli in studies with patients. Results from such recordings can provide important information on the physiological function of the nerve, on an electrode-by-electrode basis, and also can be used to evaluate the ability of speech processing strategies to produce desired spatial or temporal patterns of neural activity.

Work in this quarter included:

- Ongoing studies with Ineraid subject SR2. Studies in this quarter included (1) completion of an extensive series of measures to evaluate effects of manipulations in rate of stimulation and in the cutoff frequency for the lowpass filters in the envelope detectors in CIS processors and (2) evaluation of the TIMIT speech data base as a source of difficult sentences for sensitive measures of speech reception by a high-performance subject.
- Studies with Ineraid subject SR10, for the week beginning August 2 and August 9. The studies included (1) longitudinal measures with his portable CIS (CIS-Link) processor, (2) extension of prior studies conducted with this subject to evaluate effects of manipulations in rate of stimulation and in the cutoff frequency for the lowpass filters in the envelope detectors in CIS processors, and (3) measures of consonant identification for CIS processors using a wide range of compression functions.
- Continued development of an Access database of processor designs and study results, to bring this information together in one place for fast access and in a structure that will allow retrieval of prior designs and results on the basis of shared attributes and parameter values.
- Participation by Blake Wilson and Stefan Brill in a workshop in Frankfurt, Germany, on bilateral implants and binaural processing, at the invitation of the Med El company. (Invited speakers for the Workshop included J. Müller, F. Schön, and H. Kühn-Inacker of the Julius-Maximilians Universität in Würzburg, G. Smoorenburg of the University of Utrecht, B. Wilson of RTI, and J. Tillein of the J.W. Goethe Universität in Frankfurt. Approximately 30 people attended the workshop.)
- A visit by Wilson to the J.W. Goethe Universität in Frankfurt, at the invitation of Professor Dr. von Ilberg. Results from studies at the university to evaluate combined electric and acoustic stimulation of the same cochlea were discussed in detail, as were possibilities for future joint studies between the university and RTI to evaluate additional conditions for combined stimulation.
- A visit by Wilson to the Julius-Maximilians Universität in Würzburg, in part for further development of plans for cooperative studies between the university and RTI with recipients of bilateral COMBI 40+ implants.
- Presentation of project results in invited lectures at the *Bilateral Research Meeting* in Frankfurt and at the *30th Neural Prosthesis Workshop.*
- Continued preparation for studies with patients having bilateral COMBI 40+ implants or bilateral CI24M implants, principally by Stefan Brill, Charles Finley and consultant Marian Zerbi.
- Continued analysis of psychophysical, speech reception, and evoked potential data from current and prior studies.
- Continued preparation of manuscripts for publication.

In this report we present a detailed review of strategies developed to date for representing speech information with cochlear implants. The review provides a historical perspective for current efforts to develop better strategies and offers comments about the importance of fitting, strategy implementations, and the patient variable on performance with implants. A final section in the review mentions some possibilities for further development.

This review was originally published in the book *Cochlear Implants: Principles & Practices*, edited by J.K. Niparko, K.I. Kirk, N.K. Mellon, A. McConkey Robbins, D.L. Tucci, and B.S. Wilson (Lippincott Williams & Wilkins, Philadelphia, 2000). The reader is referred to other chapters in that book for additional information about the current status and clinical application of cochlear prostheses. Preparation of the chapter reproduced here was supported by this project (N01-DC-8-2105).

Results from the studies indicated in the bulleted list above will be presented in future reports.

7

Strategies for Representing Speech Information with Cochlear Implants

Blake S. Wilson

Remarkable progress has been made in the design and application of speech-processing strategies for cochlear implants. In particular, use of the new *continuous interleaved sampling* (CIS) and *spectral peak* (SPEAK) strategies have produced large improvements in speech reception performance compared with prior strategies (Skinner *et al.*, 1994; Wilson *et al.*, 1991a). All major manufacturers of multichannel implant systems now offer CIS or CIS-like strategies in their speech processors, with one offering both SPEAK and CIS. According to the 1995 National Institutes of Health Consensus Statement on Cochlear Implants in Adults and Children, "A majority of those individuals with the latest speech processors for their implants will score above 80-percent correct on high-context sentences, even without visual cues." Additional information on levels of performance is presented later in this chapter and in Chapter 10.

Although great progress has been made, much remains to be done. Patients with the best performance still do not hear as well as people with normal hearing, especially in adverse acoustic environments, and many patients do not enjoy high levels of performance even with the new processing strategies. The range of performance across patients is large with any of the current multichannel implant systems.

The purpose of the speech processor is to transform microphone inputs into patterns of electrical stimulation that convey the information content of speech and other sounds (see Chapter 6). This chapter describes how information is encoded in the production of speech and how such information can represented or partially represented with cochlear implants.

ELEMENTS OF SPEECH

A simple but useful model of speech production is shown in Fig. 7.1. This source-filter model (Flanagan, 1972) recognizes the first-order independence between excitation of the vocal tract and its resonant response to the excitation. Unvoiced sounds of speech are produced with a source of broadband turbulent noise. This noise is generated by forcing air through a narrow constriction (for production of unvoiced fricatives such as /s/) or by building pressure behind an obstruction and suddenly releasing the pressure with removal of the obstruction. Stop consonants, such as /t/, are produced in this way.

B. S. Wilson: Research Triangle Institute, Research Triangle Park, North Carolina 27709.

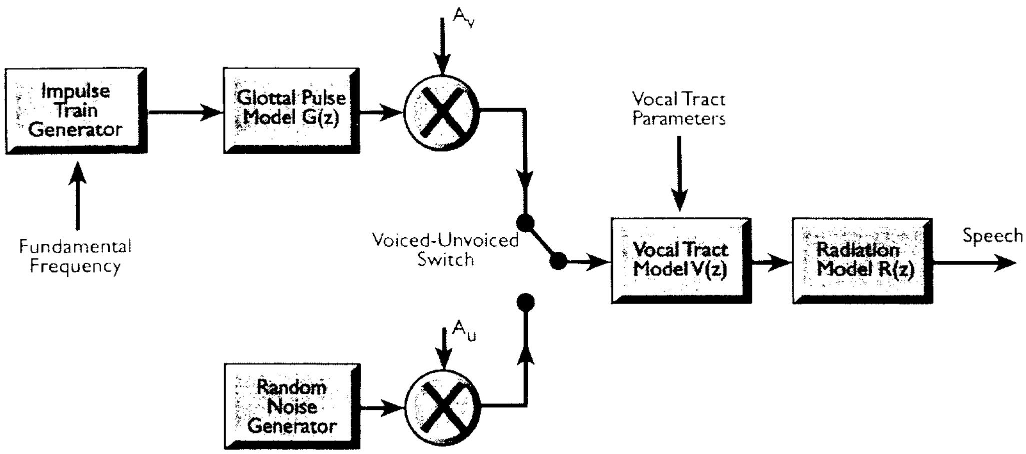

FIG. 7.1. A basic model of speech production. Parameters controlling the model include a binary indication about whether the sound is voiced or unvoiced, the frequency of glottal openings for voiced sounds, and the frequency transfer characteristics of the vocal tract. Parameters typically are updated at 5- to 30-ms intervals. The cross-marked circles indicate multiplier blocks. Speech amplitudes are controlled by the gain factors A_v and A_u for voiced and unvoiced sounds, respectively. (Adapted from O'Shaughnessy, 1987, with permission.)

The spectral characteristics of the broadband noise are altered by transmission though the vocal tract and to a lessor extent by radiation of sound at the lips.

In contrast, voiced sounds in speech are produced by exciting the vocal tract with puffs of air released through the vibrating folds of the glottis. The shape of the vocal tract, as adjusted through positions of the tongue, lip, jaw, and velum, determines how the excitation is filtered in the transmission of sound through the tract. The transfer function of the vocal tract is the *filter* in the source-filter model, and the spectral peaks in this transfer function are called *formants*. The frequencies of the first two formants convey adequate information for identification of vowels and for distinctions among other voiced sounds.

A third class of speech sounds is produced by the combination of periodic glottal excitation and aperiodic (supraglottal) noise sources. Voiced fricatives are produced in this way. For example, the addition of glottal excitation can change an /s/ sound into a /z/ sound.

Experiments with models of the type shown in Fig. 7.1 have demonstrated that a relatively small set of parameters can specify the information content of speech (Flanagan, 1972; O'Shaughnessy, 1987). In general, the parameters must specify the type of excitation (*i.e.*, voiced, unvoiced, or mixed) and the transfer function of the vocal tract. For most voiced speech sounds, the transfer function can be adequately specified by the frequencies of the first two formants. For most unvoiced speech sounds, an indication of overall spectral shape is adequate (*e.g.*, tilting up or down, principal resonance of the vocal tract). Updating such parameters every 5 to 30 milliseconds allows production of intelligible speech with a model such as the one in Fig. 7.1. The parameters specifying the transfer function of the vocal tract may be quantized along rather coarse scales and still preserve intelligibility. The information rate required to transmit these parametric data can be as low as 1,000 bits/s, which is far less than the 30,000 bits/s required for voice transmission over a typical telephone channel (Flanagan, 1972).

Taxonomy of Speech Sounds

Speech sounds can be classified according to the way in which they are produced, their

TABLE 7.1. *Classification of vowels*

Vowel	Place of constriction	Tongue height	Shape of lips	Example
i	Front	High	Spread	b*ee*t
I		High	Neutral	b*i*t
e		Upper-mid		b*ai*t
ε		Lower-mid		b*e*t
æ		Low		b*a*t
ɝ	Central, retroflex	High		b*ir*d
ə	Central	Mid		*a*bout
a		Low		French "l*a*"
u	Back	High	Rounded	b*oo*t
U		High	Slightly rounded	f*oo*t
o		Upper-mid	rounded	c*oa*t
ɔ		Lower-mid		b*ough*t
ʌ		Lower-mid	Neutral	b*u*t
ɑ		Low		h*o*t

acoustic characteristics, or the way in which they are perceived. These are interrelated; for example, the acoustic characteristics are a direct result of the way in which the sounds are produced.

Tables 7.1 and 7.2 present classifications of vowels and consonants, respectively, according to the way in which they are produced. Vowels are produced with an open vocal tract, with relatively large distances between the tongue and the roof of the mouth. Consonants are produced by narrowing or obstructing the vocal tract at some point. Air forced through narrow passages produces turbulent noise, as does the sudden release of air with the removal of an obstruction. Such sources of noise are the excitation signals for the unvoiced speech sounds.

TABLE 7.2. *Classification of consonants*

Consonant	Manner of production	Voicing	Place of constriction	Example
p	Stop	Unvoiced	Bilabial	*p*at
b		Voiced		*b*at
t		Unvoiced	Alveolar	*t*ar
d		Voiced		*d*one
k		Unvoiced	Velar	*k*ite
g		Voiced		*g*ate
f	Fricative	Unvoiced	Labiodental	*f*ar
v		Voiced		*v*iew
θ		Unvoiced	Linguadental	*th*in
ð		Voiced		*th*en
s		Unvoiced	Alveolar	*s*ee
z		Voiced		*z*oo
ʃ		Unvoiced	Palatal	*sh*ow
ʒ		Voiced		a*z*ure
h		Unvoiced	Glottal	*h*at
ʧ	Affricative	Unvoiced	Palatal	*ch*ase
ʤ		Voiced		*j*udge
m	Nasal	Voiced	Bilabial	*m*an
n			Alveolar	*n*ot
ŋ			Velar	si*ng*
l	Liquid		Alveolar	*l*oud
r			Alveolar + velar	*r*oad
w	Glide		Bilabial + velar	*w*ay
j			Palatal	*y*es

Many of the consonants also are characterized by relatively rapid movements of the articulators and consequently relatively rapid changes in the acoustic domain. Stop consonants are produced with a complete obstruction of the vocal tract followed by a sudden release of the obstruction. A noise is produced at the release, and this noise is filtered according to the shapes (principally the lengths and cross-sectional areas) of the vocal tract in front of and behind the initial obstruction.

The time between the release and onset of voicing for the following speech sound (*i.e.*, the voice onset time) is short or even negative for the voiced stop consonants, whereas the time between the release and the onset of voicing is tens of milliseconds for unvoiced stop consonants. Cognate pairs of stop consonants are identical in all respects except for the voiced or unvoiced distinction. These pairs are /p/-/b/, /t/-/d/, and /k/-/g/ in spoken English; the unvoiced members of the pairs are listed first.

Fricative consonants are produced with a narrowing of the vocal tract sufficient to generate turbulent noise. This source of noise may be accompanied by periodic puffs of air from the glottis. Unvoiced fricatives are produced with the noise source only, whereas voiced fricatives are produced with a combination of the noise source and the periodic excitation from the glottis. The voiced and unvoiced consonants also form cognate pairs: /f/-/v/, /θ/-/ð/, /s/-/z/, and /ʃ/-/ʒ/. An additional unvoiced fricative, /h/, does not have a voiced counterpart in spoken English.

Affricative consonants are concatenations of stop and fricative consonants. They can be unvoiced (/ʧ/) or voiced (/ʤ/).

Nasal consonants are produced by opening the velum passage to the nasal tract and closing the vocal tract. Different nasals are produced with different points of closure along the vocal tract. The closed vocal tract acts as a side branch resonator, which affects the spectrum of the sound emerging from the nostrils.

Liquid consonants include the lateral and retroflex consonants. The laterals (/l/ and related sounds) are produced with contact of the tip of the tongue with the roof of the mouth, although with a narrowing of the tongue to allow the passage of air around the sides of the tongue. This alters the resonant properties of the vocal tract compared with those of vowel sounds. The retroflex consonants (/r/ and related sounds) are produced with the tongue tip and tongue dorsum elevated (but not contacting the roof of the mouth), resulting in two points of constriction along the vocal tract. This also alters the resonant properties of the tract compared with the laterals and compared with vowel sounds.

The glide consonants (/w/ and /j/) are produced with closer appositions of the tongue to the roof of the mouth compared with vowels and are characterized by relatively rapid transitions compared with vowels. The appositions are not close enough to produce a separate source of turbulent noise, and all of the glides are voiced.

As indicated in Table 7.2, consonants can be classified according to manner of production and the place of constriction (*i.e.*, place of articulation). Within the broad classes of manner of production are the cognate pairs of voiced and unvoiced consonants for stops, fricatives, and affricatives. Consonants for all other manners of production are voiced.

Vowels often are classified according to the highest point of the tongue along the length of the vocal tract (*i.e.*, front to back) and with respect to the floor of the mouth (*e.g.*, low tongue position versus high tongue position). Vowels also are classified according to lip rounding, which affects the resonant properties of the vocal tract and radiation of sound at the lips.

The degree of constriction in the vocal tract ranges from open for the low and middle vowels to closed for the stops. The continuum from low to complete constriction includes low and mid vowels, high vowels, liquids and glides, fricatives, and stops. The vocal tract is closed for nasals, but the nasal tract is open.

In broad terms, the site of constriction or closure along the vocal tract affects its transfer function and resonant properties. The tract filters the excitation, which may be voiced, unvoiced, or a combination of the two.

Vocoder Theory and Models

Models of the type shown in Fig. 7.1 have been applied in analysis-synthesis or vocoder (for *vo*ice *coder*) systems for efficient transmission of speech signals. An example of such an application is the channel vocoder illustrated in Fig. 7.2. The upper panel shows a block diagram of the analysis part of the system at the transmitting end, and the lower panel shows a block diagram of the synthesis part of the system at the receiving end. Only a limited set of parameters, as extracted from the speech input in the analysis part of the system, is transmitted to the receiver. The advantage provided by analysis-synthesis systems is that the information rate required for transmission of the parameters is much less than that required for transmission of the unprocessed speech signal. Such savings allow the transmission of many more conversations through a channel of limited capacity. Extraction of parameters at the transmitting end also can allow encryption of messages for secure communications; encryption of a limited set of parameters is relatively straightforward and more secure compared with encryption of the unprocessed speech waveform.

The development of vocoder systems has a long and illustrious history, beginning with the initial development of the channel vocoder by Dudley in the late 1930s. Excellent reviews of vocoder designs and performance are presented in books by Flanagan (1972), O'Shaughnessy (1987), Papamichalis (1987), and Rabiner and Shafer (1978).

In channel vocoders, information about the excitation of the vocal tract is extracted with a voicing detector and with a pitch (or fundamental frequency) detector. The voicing detector determines whether the current speech sound is voiced or unvoiced, and the pitch detector determines the frequency of glottal openings for voiced speech sounds. Information about the configuration of the vocal tract is extracted with a bank of bandpass filters and envelope detectors. This analysis provides snapshots of the filtering by the vocal tract at 5- to 30-millisecond intervals.

The information transmitted between the analysis and synthesis ends of the system includes a binary indication of whether the sound is voiced or unvoiced, the fundamental frequency of voiced speech sounds, and the smoothed envelopes of energies within multiple bandpass ranges of the speech input. At the receiver, this information is used to reconstruct the speech waveform for the listener. The binary indication of voicing controls a switch that connects a noise source or a source of periodic pulses to the inputs of a set of multiplier blocks. The rate of the periodic pulses for voiced speech sounds is controlled by the parameter specifying the frequency of glottal openings. The other inputs to the multiplier blocks are the envelope signals for each of the bandpass channels. The outputs of the multiplier blocks are directed to a bank of bandpass filters (corresponding to the bank in the analysis portion of the system). A synthesized speech signal is formed by summing the outputs of the bandpass filters.

The performance of channel vocoders is affected by the accuracy of parameter extraction and by many choices in design, including the resolution with which the channel and fundamental frequency parameters are quantized and the number and frequency boundaries of the bandpass filters. In general, performance is improved with increases in resolution and increases in the number of bandpass filters up to points corresponding to limits in the *perception* of speech and other sounds. For example, multiple bandpass filters within critical bands of hearing are no better than one bandpass filter for each of the critical bands. Depending on the selected endpoints, 14 to 19 critical bands span the range of speech frequencies. This has led to

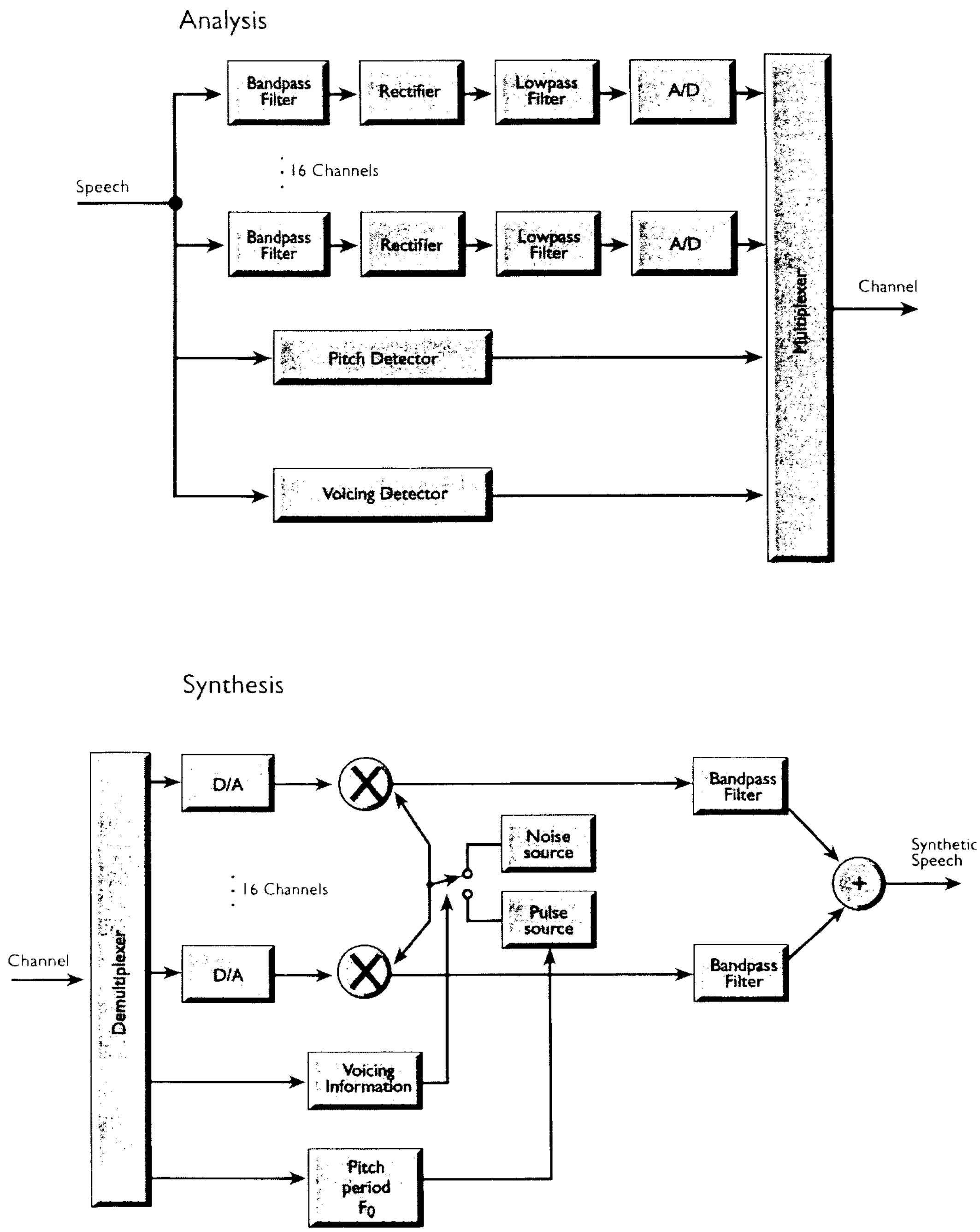

FIG. 7.2. Diagrams of the analysis *(top)* and synthesis *(bottom)* portions of a channel vocoder. The blocks labeled A/D *(top)* correspond to analog-to-digital converters, and the blocks labeled D/A *(bottom)* correspond to digital-to-analog converters. The cross-marked circles indicate multiplier blocks, and the circle with a plus mark indicates a summation block. (Adapted from Papamichalis, 1987, with permission.)

the design of channel vocoder systems with 14 to 19 bandpass channels, with the frequency boundaries for the channels distributed along a logarithmic scale corresponding to the distribution of frequency boundaries for critical bands in hearing. Departures from the logarithmic scale or reductions in the number of channels usually produce decrements in performance.

Synthesized speech produced by a channel vocoder is not perfectly intelligible even with a high number of bandpass channels and fine quantization of the transmitted parameters. One limitation of channel vocoders is in the binary decision between voiced and unvoiced sounds. Obviously, the binary decision does not recognize mixed excitations of the vocal tract, such as those of the voiced fricatives. This oversimplification eliminates some important distinctions among speech sounds. Also, reliable determination of whether a sound is predominately voiced or unvoiced and reliable determination of the fundamental frequency for voiced speech sounds are difficult speech analysis problems, especially for speech in typical acoustic environments with reverberation and competing speakers or other noise. Unavoidable errors in the extraction of these parameters can produce further decrements in performance.

An improvement in the quality and intelligibility of transmitted speech can be realized with a different approach for representing the excitation of the vocal tract. In this approach, the speech input is filtered to include frequencies below a cutoff frequency in the range of 400 to 1,000 Hz. This filtered or baseband signal then is transmitted without further analysis to the receiver. At the receiver, the baseband signal is used for the source of excitation. No decisions are made about voicing and no analysis is required to determine fundamental frequencies for voiced sounds. Instead, the baseband signal reflects the excitation of the vocal tract. It is periodic for voiced speech sounds, aperiodic for unvoiced speech sounds, and has periodic and aperiodic components for mixed excitation of the vocal tract. The periodicity of the voiced sounds corresponds to the rate of glottal openings.

The bandpass channels of this baseband vocoder are the same as those used in the channel vocoder. Transmission of the baseband signal requires a much higher information rate than that required for the transmission of the voiced or unvoiced and fundamental frequency parameters in the channel vocoder. There is a tradeoff in quality versus information rate requirements in choosing between channel and baseband vocoders. Baseband vocoders provide higher quality, especially in noisy environments, but at the cost of higher information rate requirements compared with channel vocoders.

The representation of how the vocal tract is excited is far from perfect, even in baseband vocoders. In particular, only a small part of the broad spectrum of noise excitation is represented in the baseband signal, and this can lead to errors in levels of excitation for unvoiced sounds.

Another approach to vocoder design is to extract and transmit parameters to specify the principal resonances and anti-resonaces of the vocal tract, as opposed to the transmission of bandpass envelope signals in the channel vocoder. This approach is based on the observation that peaks in the spectra of voiced speech sounds convey almost as much information as the full spectrum. For vowels and some of the other voiced sounds, a savings in the information rate required for transmission can be realized by sending a reduced set of parameters that specify the frequencies and widths of the first two to four peaks (formants) in speech spectra at 5- to 30-millisecond intervals. A further savings can be realized by specifying the frequencies only, because the widths convey less information than the frequencies and can be fixed for demanding applications.

A block diagram of the synthesis portion of such a formant vocoder is presented in Fig. 7.3. The upper path is used for the synthesis of voiced speech sounds, and the lower path is used for the synthesis of unvoiced speech sounds. The parameters F_1 through F_3 specify the center frequencies of three resonant filters connected in series, and the parameters B_1 through B_3 specify the widths (or bandwidths) of the filters.

Synthesis of unvoiced speech sounds requires an anti-resonance (spectral zero) in addition to a resonance (spectral pole), corresponding the cavities between the source of noise and the lips (resonance) and between the source of noise and the glottis

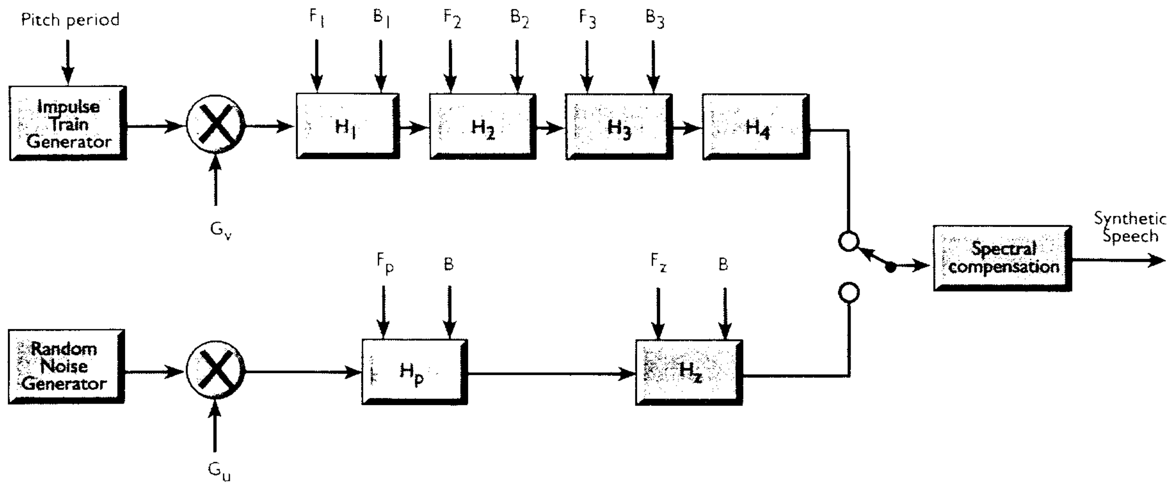

FIG. 7.3. Diagram of the synthesis portion of a formant vocoder. Blocks labeled H_1 to H_3 are resonance filters with center frequencies controlled by the parameters F_1 to F_3 and with bandwidths controlled by the parameters B_1 to B_3. The filters represent the first three formants of voiced speech sounds. Block H_4 is a resonance filter whose parameters are fixed and do not vary with time. This filter represents the relatively invariant fourth formant of voiced speech sounds. Blocks H_p and H_z represent the principal resonance (*i.e.*, spectral pole) and anti-resonance (*i.e.*, spectral zero) of unvoiced speech sounds, respectively. Parameters F_p and F_z control the center frequencies of those filters, and the single parameter B controls the bandwidth for both of the filters. Speech amplitudes are controlled by the gain factors G_v and G_u for voiced and unvoiced sounds, respectively. (Adapted from Papamichalis, 1987, with permission.)

(anti-resonance). The frequencies of the resonance and anti-resonance are specified by the parameters F_p (pole) and F_z (zero), respectively. The selectivity for both filters is specified by the single parameter B.

The formant vocoder offers a reduction in the information rate needed to transmit the extracted parameters. However, this reduction comes at the cost of a great increase in the complexity of the analysis part of the system. Accurate extraction of formant frequencies (and bandwidths) in conversational speech is difficult and highly subject to noise interference.

The formant vocoder also shares some limitations of the channel vocoder previously discussed. In particular, the weaknesses associated with the voiced or unvoiced decision and with extraction of the fundamental frequency for voiced speech sounds apply to both types of vocoder. The formant vocoder also fails to represent nasals well, in that accurate synthesis of nasals requires an anti-resonance (produced by the closed cavity of the vocal tract) in addition to the resonances of other voiced sounds. The anti-resonance is not included in the path for the synthesis of voiced sounds.

Formant vocoders work best for transmission and synthesis of vowels and work less well for transmission and synthesis of other speech sounds. Formant vocoders also are more complex and more susceptible to noise interference than channel vocoders. Formant vocoders have been used in situations requiring extremely low information rates for transmission of speech parameters. The quality and intelligibility of speech transmitted with formant vocoders is not as high as those with channel vocoders, but the formant vocoder allows lower information rates while maintaining an acceptable intelligibility for some applications.

Implications for Cochlear Implants

The amount of information that can be presented and perceived with a cochlear implant is much less than that for someone with normal hearing listening to an unprocessed acoustic signal. For example, the number of electrodes that can be included in a scala tympani implant is far less than the number of ganglion cells and neurons in normal and probably in most deafened cochleas. The rate at which stimulus information can be sent to the electrodes is further restricted by the properties of transcutaneous links (see Chapter 6). Current implant systems have no more than 22 intracochlear electrodes, which for such closely spaced electrodes address highly overlapping populations of neurons because of the spread of the electric fields across electrode positions. Data from studies manipulating the number of channels and electrodes used with the 22-electrode implant indicate that addition of channels and electrodes beyond four to six does not increase speech reception scores (Fishman *et al.*, 1997; Lawson *et al.*, 1996; Wilson, 1997). This suggests that only four to six sites along the array elicit clearly discriminable percepts. This number may be increased with the use of new electrode designs (see Chapter 6). This number is lower than the number of critical bands in hearing that span the range of speech frequencies (14 to 19 critical bands), much lower than the number of rows of sensory hair cells (approximately 3,000, which probably correspond to the smallest units of frequency resolution according to place of stimulation in the cochlea), and very much lower than the number of auditory neurons (approximately 30,000 in the healthy cochlea).

In addition to the limitations of the implant system, perception of electrical stimuli is different from perception of acoustic stimuli. The dynamic range of stimulus amplitudes from auditory threshold to loud percepts is on the order of 10 to 20 dB for electrical pulses and on the order of 100 dB for acoustical stimuli. The number of discriminable steps in stimulus amplitudes within these ranges is lower in electrical hearing than in normal hearing. Changes in the rate or frequency of stimuli delivered to single electrodes of an implant are perceived as changes in pitch only up to a pitch saturation limit, typically around 300 pulses/s for electrical pulses or 300 Hz for electrical sinusoids. Higher rates or frequencies do not produce increases in pitch. In normal hearing, different pitches are heard over much wider ranges of rates or frequencies, probably through combinations of rate and place cues to pitch.

A wide range of outcomes is found for any of the current multichannel implants (see Chapter 10). Different patients using identical implant devices may have quite different speech reception scores. This indicates the importance of patient variables in the design and performance of implant systems. Such variables may include differences among patients in the survival of neural elements in the implanted cochlea, proximity of the electrodes to the target neurons, depth of insertion for the electrode array, integrity of the central auditory pathways, and cognitive and language skills.

The various limitations described above suggest a highly impoverished link for the representation and perception of speech information with cochlear implants. Indeed, design of implant systems can be viewed as a problem of squeezing speech through a narrow bottleneck, imposed by the lack of spatial specificity in stimulation and by limitations in perception. Early designs attempted to represent only the most important aspects of speech at the electrodes, and they attempted to match the representations with what could be perceived. This involved transformations of extracted parameters such that the minimum and maximum values of each parameter would span a perceptual dimension, such as from auditory threshold to loud percepts.

Vocoder theory and models played major roles in these early designs, at least for multi-electrode implant systems. Vocoder results demonstrated that a small set of parameters could specify the synthesis of intelligible speech at low rates of information transfer. The results also indicated how such parameters could be extracted from continuous speech.

PROCESSING STRATEGIES

The development of cochlear prostheses began with the work of Djourno and Eyries in 1957 and with the separate efforts led by House, Simmons, and Michelson in the 1960s. Many approaches to the design of processing strategies and other components such as electrode arrays have been proposed or used in the 4 decades of work from these beginnings to the present.

Only a few among the many approaches to processor design are described in this chapter. Emphasis is given to the most recent strategies, and to the major steps that led to the development of those strategies. Emphasis also is given to within-subject comparisons of processing strategies, in which effects of processor variables can be separated from effects of patient variables (Wilson *et al.*, 1993). These comparisons include CIS versus a *compressed analog* (CA) strategy and SPEAK versus a *multipeak* (MPEAK) strategy. The comparisons illustrate issues in processor design and show what is possible with the use of the CIS and SPEAK strategies. Descriptions of many more of the approaches to processor design that have been proposed or used throughout the history of cochlear implants are presented in a number of comprehensive reviews (Dorman, 1993; Gantz, 1987; Loizou, 1998; Millar *et al.*, 1984 and 1990; Moore, 1985; Parkins, 1986; Pfingst, 1986; Tyler and Tye-Murray, 1991; Wilson, 1993).

Interleaved Pulses Strategies

Early designs of processing strategies based on vocoder analogies included the *interleaved pulses* (IP) strategies and a series of feature extraction strategies. An approach similar to that of channel vocoders was used in the IP strategies, and an approach similar to that of formant vocoders was used in the feature extraction strategies.

A block diagram of the IP processor (Wilson *et al.*, 1985 and 1988a) is presented in Fig. 7.4. The "front end" of the processor is nearly identical at the block diagram level to the analysis portion of a channel vocoder. The front end includes a bank of bandpass filters and associated envelope detectors, a voicing detector. The outputs from these blocks then are used to control patterns of electrical stimulation at the electrode array.

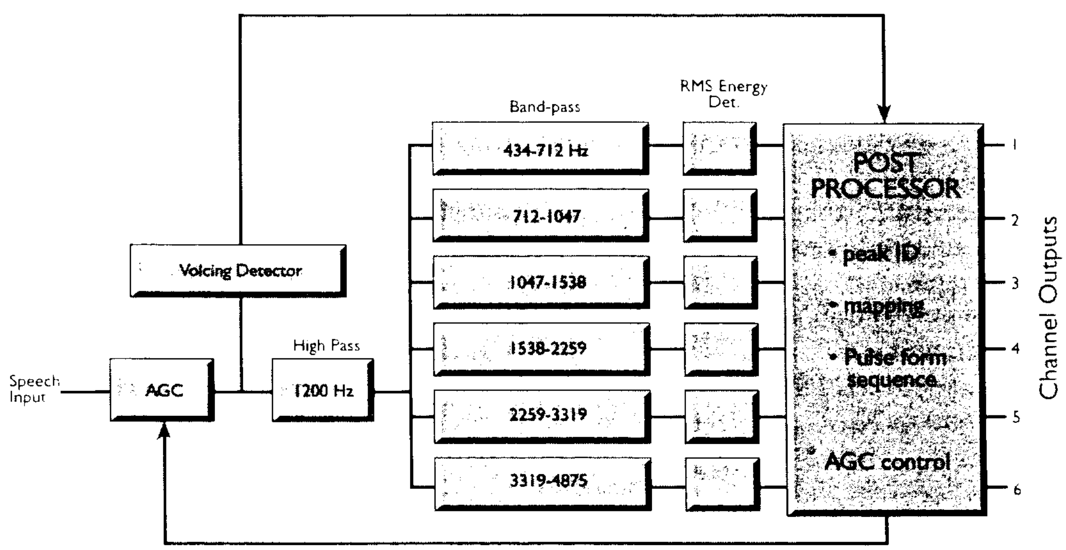

FIG. 7.4. Diagram of an interleaved pulses processor. The initial stages of processing are performed by a voicing detector and a bank of bandpass filters and associated envelope detectors. A postprocessor uses inputs from these stages to control patterns of stimulation at the electrode array. AGC, automatic gain control. (Adapted from Wilson *et al.*, 1988a, with permission.)

A *postprocessor* determines the timing and amplitudes of stimulus pulses delivered to the electrodes. In variation 1 of the IP processor, it presents pulses at the fundamental frequency when the voicing detector indicates the presence of a voiced speech sound, or at a relatively high (*e.g.*, 300 pulses/s/electrode) or randomized rate when the voicing detector indicates the presence of an unvoiced speech sound.

In variation 2 of the IP processor, inputs from the voicing and fundamental frequency detectors are ignored and pulses are presented at fixed interpulse intervals (*e.g.*, 1 millisecond) to a subset of selected electrodes. The electrodes are selected according to the *n* highest envelope signals among the *m* bandpass channels for each cycle of stimulation. This variation of the IP processor was the first implementation of an *n*-of-*m* strategy for cochlear implants (Wilson *et al.*, 1985 and 1988; McDermott and Vandali, 1997). The basic *n*-of-*m* approach is used to this day, as described later in this chapter.

Patterns of stimulation for variation 1 of the IP processor are illustrated in Fig. 7.5. Speech inputs are shown in the top panel, and stimulus pulses are shown for each of four electrodes (and channels) in the bottom panel. The four electrodes are arranged in an apex-to-base order, with electrode 4 being most basal. The amplitudes of the pulses for each of the electrodes are derived from the envelope signals in the corresponding bandpass channels. The envelope signal in the bandpass channel with the lowest center frequency controls the amplitudes of pulses delivered to the apical-most electrode, and the envelope signal in the bandpass channel with the highest center frequency controls the amplitudes of pulses delivered to the basal-most electrode. This arrangement mimics the tonotopic organization of the cochlea in normal hearing, with high-frequency sounds exciting neurons at basal locations and low-frequency sounds exciting neurons at apical locations.

Pulse amplitudes are derived from the envelope signals in both variations of IP processors. A logarithmic transformation is used to map the relatively wide dynamic range of envelope variations onto the narrow dynamic range of electrically evoked hearing. (The effect of such mapping is not shown in Fig. 7.5.)

Variation 1 of the IP processors is a close analog of a channel vocoder. As illustrated in Fig. 7.5, the stimuli produced with this variation indicate voiced versus unvoiced sounds (compare rates of stimulation between the lower panels of the figure, corresponding to the voiced /ɔ/ and the unvoiced /t/) and the fundamental frequency of voiced sounds (*i.e.*, the pulse rate in the lower left panel). Variations in pulse amplitudes across electrodes also indicate the shape and transfer function of the vocal tract; at least some of the principal resonances and anti-resonances of the tract are reflected in the different pulse amplitudes for different electrodes. For example, the high-frequency resonance of the /t/ sound is reflected in the relatively intense stimulation of electrode 4 (lower right panel of Fig. 7.5).

Variation 2 of the IP processors uses the bandpass channel outputs only and represents only a subset of those for each cycle of stimulation across electrodes. The presentation of information is sparse compared with variation 1 of the IP processors. Such a presentation may provide a more optimal match for patients with highly limited perceptual abilities (Wilson *et al.*, 1988a). Information about the excitation of the vocal tract is discarded, but some information about the transfer function of the tract is retained.

An additional aspect of the design of IP processors is the nonsimultaneous presentation of pulses across electrodes. This eliminates direct summation of electric fields from different electrodes produced with simultaneous stimulation. Such summation can greatly reduce the independence of stimulation among electrodes and thereby reduce the salience of channel-related cues.

Comparisons between IP and other processing strategies are described in several reports (Lawson *et al.*, 1993; Wilson, 1993; Wil-

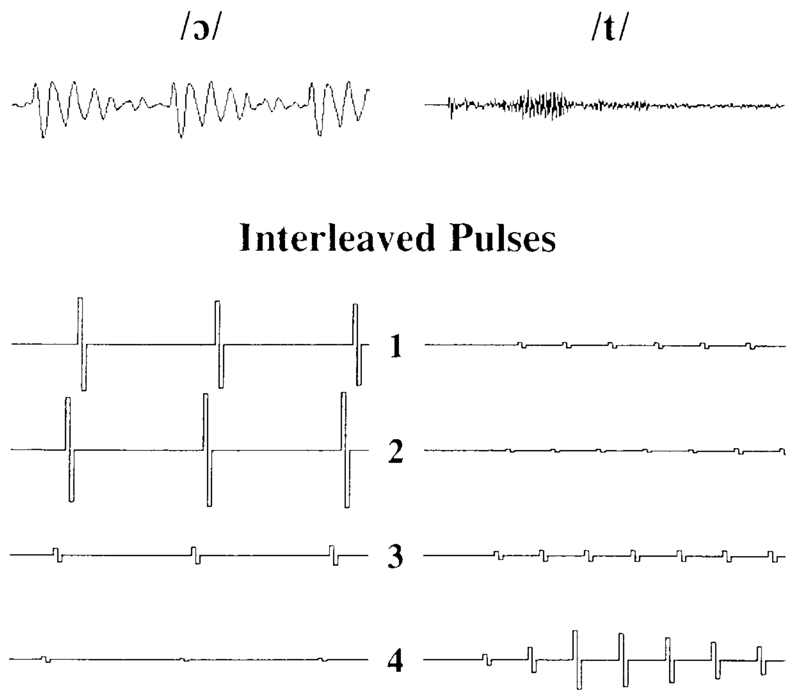

FIG. 7.5. Stimuli produced by a simplified implementation of variation 1 of an interleaved pulses (IP) processor. The top panels show pre-emphasized (6-dB/octave attenuation below 1.2 kHz) speech inputs. An input corresponding to a voiced speech sound (/ɔ/) is shown in the left panel, and an input corresponding to an unvoiced speech sound (/t/) is shown in the right panel. The remaining panels show stimulus pulses produced by an IP processor for these inputs. The numbers indicate the electrodes to which the stimuli are delivered. The lowest number corresponds to the apical-most electrode and the highest number to the basal-most electrode. The pulse amplitudes in the figure reflect the amplitudes of the envelope signals for each channel. In actual implementations, the range of pulse amplitudes is compressed using a logarithmic or power-law transformation of the envelope signal for each channel. Notice that the stimuli are presented nonsimultaneously across electrodes and that the rates of stimulation are different for voiced and for unvoiced segments of speech. The duration of each trace in this figure is 25.4 ms. (From Wilson, 1993, with permission.)

son *et al.*, 1985, 1988a, 1988b, and 1991b). Performance for one subject was remarkably improved with substitution of variation 2 of the IP processors for the CA processor of his clinical device. This subject did not recognize speech as such with the clinical processor. In contrast, speech sounded like speech with the IP processor, and the subject scored well above chance on tests of consonant and vowel identification. The amount of information presented with the IP processor was much less than that presented with the CA processor (discussed later in this chapter), and this may have helped the subject to perceive a limited but important subset of speech information. The nonsimultaneous presentation of stimuli with the IP processor, as opposed to the simultaneous presentation of analog waveforms with the CA processor, may have enhanced the representation and perception of channel-related cues.

This second variation of IP processors has been developed further in *n*-of-*m* strategies in current use. These designs are described later in this chapter.

Additional comparisons involving IP processors included comparisons of the first variation of the processors with CA processors in groups of six subjects using the USCF/Storz implant system and two subjects using the Ineraid implant system. All of these subjects had higher levels of speech reception performance with their clinical CA processors compared with the subject described previously. In general, results obtained with IP processors with these groups of subjects were immediately as good on average as results obtained with the CA processor, despite considerable experience with the latter and essentially no experience with the former. Some subjects had better performance with the CA processor, and others had better performance with the IP processor. This finding suggested that availability of multiple strategies might help individual patients to achieve the best possible outcome.

Although the results with variation 1 of the IP processor were promising, its further development was discontinued in favor of a clearly better approach, the CIS strategy, that also was evaluated with the second group of two Ineraid subjects.

Feature Extraction and Multipeak Strategies

Several processing strategies have been used in conjunction the Nucleus electrode array and implanted receivers since 1982, when the first system was introduced for clinical application (Clark, 1987; Clark *et al.*, 1990; Patrick and Clark, 1991). The first three of the strategies were based largely or in part on a formant vocoder model. The SPEAK, CIS, and *advanced combination encoder* (ACE) strategies have superseded these earlier strategies and are the processing options now available for use with the Nucleus implant. These strategies are described later in this chapter.

The first strategy used with the Nucleus implant was designed to represent voicing information and the frequency and amplitude of the second formant (F2 and A2, respectively). The second strategy added a representation of the first formant (F1 and A1).

Block diagrams of the second strategy are presented in Figs. 7.6 and 7.7. Fig. 7.6 shows a diagram of the entire processor, and Fig. 7.7 shows a more detailed diagram of the front end of the processor. The first processor in the Nucleus series used the components associated with extraction and representation of voicing and second formant information, and the second processor used all of the components shown in Figs. 7.6 and 7.7, including those associated the extraction and representation of first formant information.

In the first strategy, a zero crossings detector was used to estimate the fundamental frequency (F0) of voiced speech sounds from the output of a 270-Hz lowpass filter (Fig. 7.7, upper path). A separate zero crossings detector was used to estimate the spectral centroid (or spectral center of gravity) in the output of a bandpass filter spanning the frequency range of the second formant (1 to 4 kHz in one implementation of the strategy). The amplitude of the second formant was

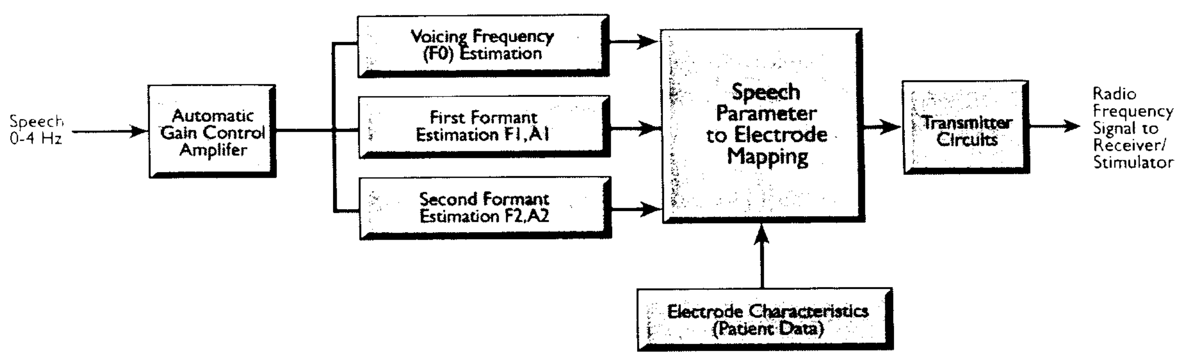

FIG. 7.6. Diagram of the F0/F1/F2 processing strategy. The F0/F2 strategy uses all blocks shown here except for the block labeled First Formant Estimation F1,A1. (Adapted from Patrick *et al.*, 1990, with permission.)

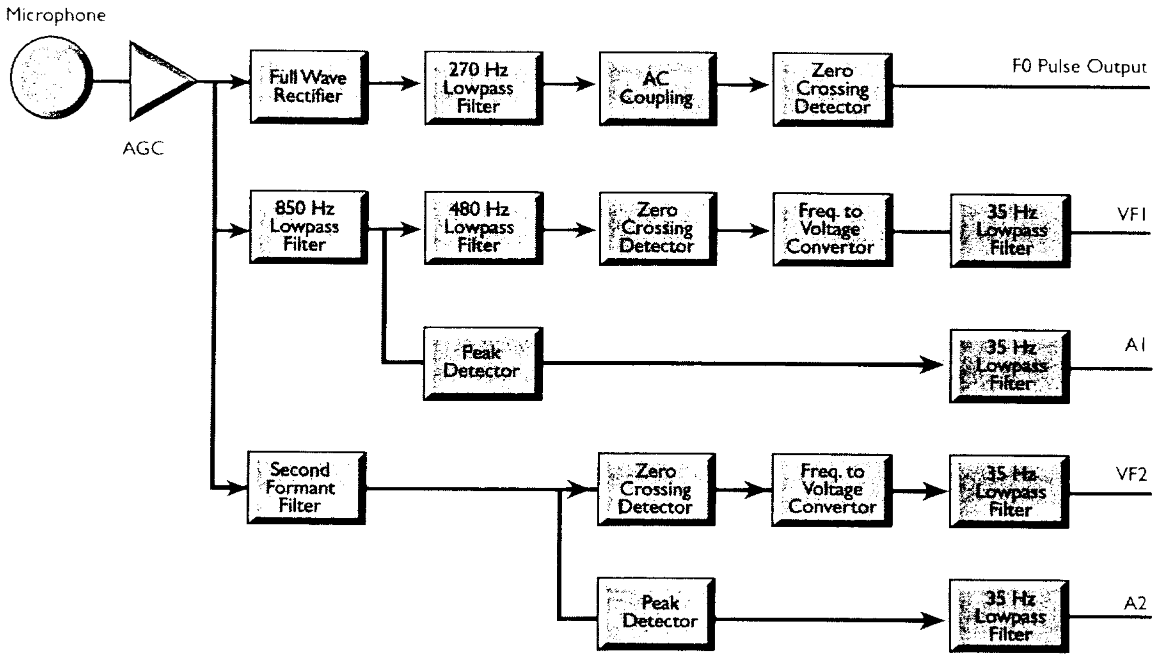

FIG. 7.7. Detailed diagram of the front end of an F0/F1/F2 processor. AGC, automatic gain control. (From Blamey *et al.*, 1987, with permission.)

estimated with an envelope detector at the output of the bandpass filter. The envelope detector included a peak detector and lowpass filter, whose cutoff frequency was set at 35 Hz. These components are shown in the lower path of Fig. 7.7.

This F0/F2 processor represented voicing information by presenting stimulus pulses at the estimated F0 rate during voiced speech sounds and at quasi-random intervals, with an average rate of about 100 pulses/s, during unvoiced speech sounds. The frequency of the spectral centroid in the F2 band was represented by selecting one position along the electrode array for each successive pulse. High frequencies were represented with stimulation of an electrode or a pair of closely-spaced bipolar electrodes near the basal end of the array, and low frequencies were represented with stimulation of an electrode or electrodes near the apical end of the array. The estimated amplitude of the second formant was represented with the amplitude of each pulse, derived with a logarithmic transformation of the envelope signal, as in the IP processors described previously.

The F0/F2 strategy is a highly reduced analog of a formant vocoder. Only one formant is represented. However, this formant probably confers the greatest information about speech among the formants. Perception of F2 can allow the classification of vowels into groups, and transitions of F2 from vowels into consonants can cue place of articulation for some of the consonants. The aim of the F0/F2 strategy is to extract from speech an irreducible set of parameters and then to represent those parameters in a way that can be perceived and used by implant patients. Performance with this initial strategy was encouraging in that its use allowed some patients to recognize portions of speech with hearing alone (Clark, 1987 and 1995; Clark *et al.*, 1990; Dowell *et al.*, 1986 and 1987a). The average of scores among 13 patients for recognition of monosyllabic Northwestern University Auditory Test 6 (NU-6) words was 4.9% correct (Dowell *et al.*, 1987a).

In late 1985, the F0/F2 strategy was modified to include extraction and representation of information about the first formant. An additional channel of processing was used to derive estimates of the spectral centroid and amplitude in a band of frequencies encompassing the range of F1 (Fig. 7.7, middle path). For each stimulus cycle, a postprocessor (Fig. 7.6) selected two electrode positions for stimulation, one corresponding to the estimated frequency of F1 and the other corresponding to the estimated frequency of F2. As in the F0/F2 processor, the electrodes were stimulated at a rate equal to the estimated F0 during voiced speech and at a quasi-random rate during unvoiced speech. A stimulus cycle included stimulation of the electrode or electrode pair selected to represent F2, followed by stimulation of the electrode or electrode pair selected to represent F1.

Within-subject comparisons of the F0/F2 and F0/F1/F2 strategies demonstrated higher levels of speech reception with the latter (Dowell *et al.*, 1987b; Tye-Murray *et al.*, 1990). All seven subjects in the study of Dowell *et al.*, for instance, obtained higher scores for recognition of key words in the Central Institute for the Deaf (CID) sentences of everyday speech using the F0/F1/F2 processor. The average score for these live-voice presentations of the sentences was 30.4% correct for the F0/F2 processor and 62.9% correct for the F0/F1/F2 processor. The average scores for separate groups of subjects in another study (Dowell *et al.*, 1987a) were 15.9% correct for the F0/F1 processor group (n = 13) and 35.4% correct for the F0/F1/F2 processor group (n = 9). Scores for recognition of monosyllabic words in this study was 4.9% correct for the F0/F2 group, and 12.4% correct for the F0/F1/F2 group. Test items were presented from recorded material in the second study, and this may have contributed to the lower scores for the CID sentence tests in that study compared with the first study. Scores with the F0/F1/F2 processor were significantly higher than scores with the F0/F2 processor in both studies.

Although performance with the F0/F1/F2 strategy was better than that with the F0/F2 strategy, patients using the F0/F2 strategy for an extended period in their daily lives initially rejected the F0/F1/F2 strategy as inferior (Dowell *et al.*, 1987b). Ultimately, the patients preferred and performed better with the F0/F1/F2 strategy. This, along with similar observations by others, suggests that preference is an unreliable guide at best in selecting processing strategies for implants. In addition, a substantial period of experience or learning may be required before asymptotic performance is attained with a new strategy (Dorman and Loizou, 1997; Pelizzone *et al.*, 1995; Tyler *et al.*, 1986).

From 1985 to 1989, Cochlear Ltd. (then a subsidiary of Nucleus Ltd.), in collaboration with investigators at the University of Melbourne, developed new external hardware for use with the Nucleus implant (Patrick and Clark, 1991; Patrick *et al.*, 1990; Skinner *et al.*, 1991). Analog components were replaced with digital components, and various aspects of the signal processing and mapping of envelope levels onto stimulus levels were refined in the new hardware (Skinner *et al.*, 1991; Wilson, 1993). Use of a custom integrated circuit for much of the processing allowed substantial reductions in the size and weight of the new Mini Speech Processor (MSP) compared with the prior Wearable Speech Processor (WSP III). These reductions helped to make the MSP more suitable for use by young children.

The MSP could be programmed to implement versions of the F0/F2 and F0/F1/F2 strategies, with the refinements provided by the MSP hardware. In addition, a new strategy, MPEAK, could be implemented through software choices. The MPEAK strategy was designed to augment the F0/F1/F2 strategy by adding a representation of envelope variations in high-frequency bands of the input speech signal.

As described previously, a formant vocoder approach cannot provide a good representation of many speech sounds, particularly unvoiced sounds and many of the voiced consonants. The high-band channels of the MPEAK strategy are similar in design to the upper channels of the IP processors, which are based on a channel vocoder approach. The MPEAK strategy combines aspects and channel and formant vocoders, with the aim of improving the representation and perception of consonants.

A block diagram of the MPEAK strategy is presented in Fig. 7.8. This diagram is in the style of Fig. 7.6 for the F0/F2 and F0/F1/F2 strategies and shows all components of a MPEAK processor. The principal difference between the F0/F1/F2 and MPEAK strategies is in the addition of the energy indicators for three high-frequency bands. The bands are 2.0 to 2.8 kHz (band 3, as distinguished from the two bands for F1 and F2), 2.8 to 4.0 kHz (band 4), and 4.0 to 7.0 kHz (band 5).

In the MPEAK strategy, four pulses are delivered in each stimulus cycle. During voiced speech, these cycles are presented at a rate equal to the estimated F0 and, during unvoiced speech, at quasi-random intervals but with an average rate in the range of 200 to 300 pulses/s. Fixed electrode positions are reserved at the basal end of the array for representations of the envelope signals in bands 3 through 5, and the remaining (more apical) electrode positions are used for representations of F1 and F2. During voiced speech, the electrodes for bands 4 and 3 and for F2 and F1 are selected for stimulation. During unvoiced speech, the electrodes for bands 5, 4, and 3 and for F2 are selected. Stimulation of the four electrodes (or electrode pairs) selected for each cycle is in a base-to-apex order. Manipulations of pulse amplitude and pulse duration are used to code loudness. (These manipulations produce changes in the charge per phase of stimulus pulses, which is strongly correlated with the loudness of auditory percepts.) Changes in either can be used to code loudness, but the combination allows a reduction in the time required for the transmission of stimulus information through the transcutaneous link of the Nucleus device. In particular, transmission of high-amplitude, short-duration pulses requires less time than transmission of low-amplitude, long-duration pulses (Crosby *et al.*, 1985; Shannon *et al.*, 1990).

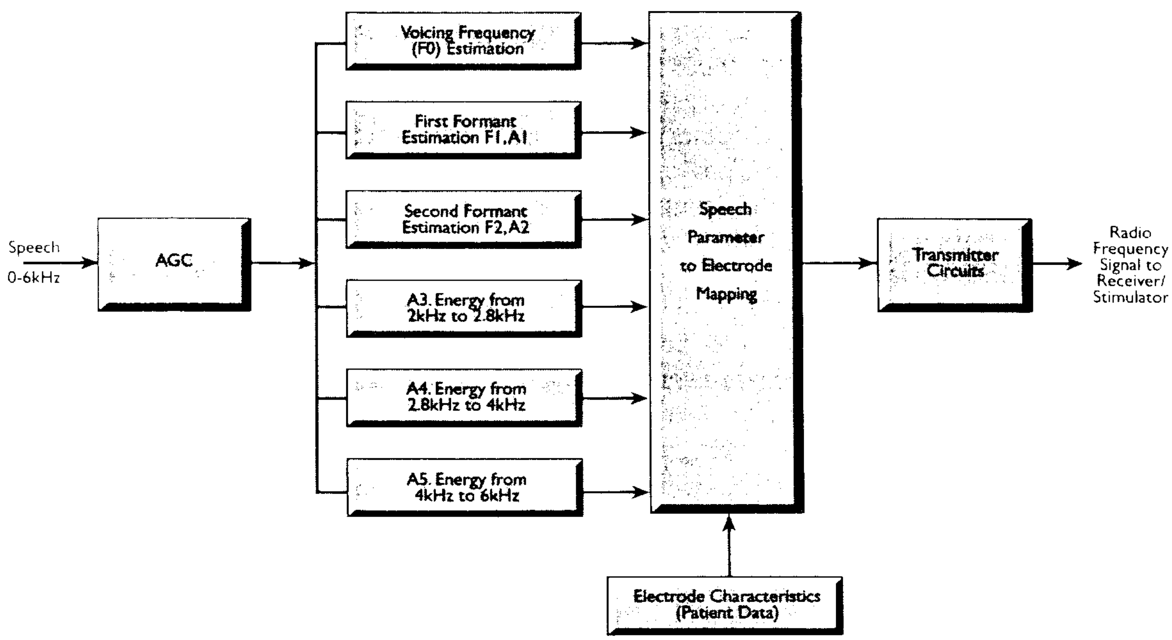

FIG. 7.8. Diagram of the multipeak processing strategy and the hardware components of the Mini Speech Processor (MSP). AGC, automatic gain control. (Adapted from Patrick *et al.*, 1990, with permission.)

Coding loudness with changes in the amplitude of short-duration pulses up to the amplitude (current) limit of the device and then coding further increases in loudness with increases in duration at that highest amplitude minimize the amount of time required to transmit each stimulus pulse. This allows rates of stimulation up to about 400 cycles/s.

Studies have been conducted to compare the F0/F1/F2 strategy as implemented in the WSP III, the F0/F1/F2 strategy as implemented in the MSP or a prototype of the MSP, and the MPEAK strategy as implemented in the MSP (Dowell *et al.*, 1991; Skinner *et al.*, 1991). In within-subject comparisons with five subjects, Dowell *et al.* found significant increases in the recognition of key words in the Bamford-Kowal-Bench (BKB) sentences when the MSP implementation of the F0/F1/F2 strategy was substituted for the WSP III implementation of that strategy, and when the MPEAK strategy was substituted for the F0/F1/F2 strategy in the MSP hardware. The scores increased from approximately 50% correct with the WSP III implementation of the F0/F1/F2 strategy to approximately 80% correct with MPEAK strategy (using the MSP hardware). Results for the MSP implementation of the F0/F1/F2 strategy were almost exactly midway between these.

Skinner *et al.* also observed large improvements in speech reception scores when the MPEAK strategy was substituted for either implementation of the F0/F1/F2 strategy in studies with a separate group of five subjects. Scores for the two implementations of the F0/F1/F2 strategy were not statistically different, however. The MSP implementation used a prototype of the MSP with some differences in design, and those differences may have affected the comparisons. Average scores for recognition of NU-6 monosyllabic words improved from 13.3% to 29.1% correct when the MPEAK strategy was substituted for the WSP III implementation of the F0/F1/F2 strategy, and recognition of key words in the BKB sentences improved from 51.0% to 70.0% correct for the same processor conditions.

Continuous Interleaved Sampling Strategy

The CIS strategy was the direct descendant of the IP strategies described earlier. The design of the CIS strategy was motivated in part by a desire to represent voicing information in a more natural way and to increase the amount of information transmitted and perceived with the implant. With the first variation of the IP processing, in which voicing information was explicitly represented, subjects remarked that changes in their percepts at boundaries between voiced and unvoiced sounds seemed abrupt and unnatural. The chunkiness of the representation did not reflect the smoother transitions in speech and did not reflect mixed excitations of the vocal tract.

These anecdotal comments by subjects were corroborated by voicing errors in tests of consonant identification. The errors may have been produced by an incomplete and distorted representation of voicing information and by errors in the voiced or unvoiced decision. The former errors are inherent to channel and formant vocoder approaches, and the latter errors are difficult or impossible to eliminate for less than ideal speech inputs, even using highly sophisticated algorithms and hardware (Hess, 1983).

A block diagram of the CIS strategy is presented in Fig. 7.9. Inputs from a microphone and optional automatic gain control (AGC) are directed to a pre-emphasis filter, which attenuates frequency components below 1.2 kHz at 6 dB/octave. This pre-emphasis helps relatively weak consonants (with a predominant frequency content above 1.2 kHz) compete with vowels, which are intense compared with most consonants and have strong components below 1.2 kHz.

The output of the pre-emphasis filter is directed to a bank of bandpass channels. Each channel includes stages of bandpass filtering, envelope detection, and compression. Envelope detection is accomplished with a

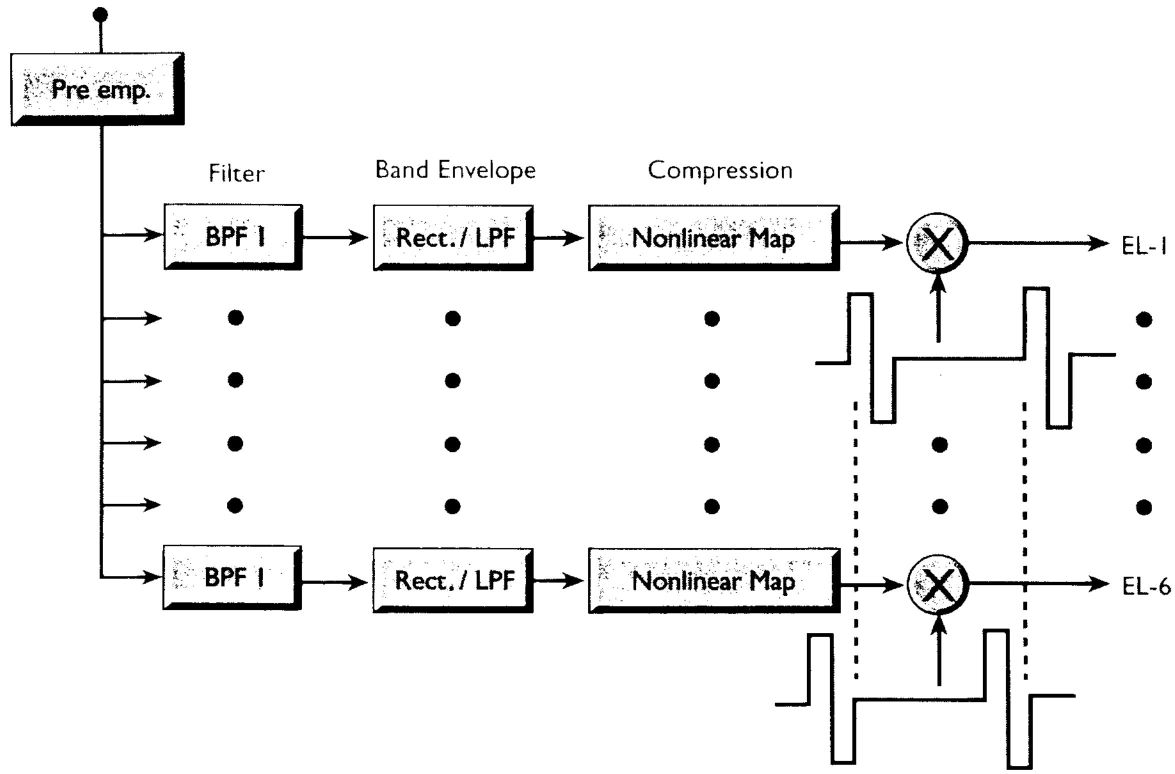

FIG. 7.9. Diagram of continuous interleaved sampling (CIS) processing strategy. The strategy uses a pre-emphasis filter (Preemp) to attenuate strong components in speech below 1.2 kHz. The pre-emphasis filter is followed by multiple channels of processing. Each channel includes stages of bandpass filtering (BPF), envelope detection, compression, and modulation. Carrier waveforms for two of the modulators are shown below the two corresponding multiplier blocks. (Adapted from Wilson *et al.*, 1991a, with permission.)

rectifier, followed by a lowpass filter. The channel outputs are used to modulate trains of biphasic pulses. The modulated pulses for each channel are applied through a percutaneous or transcutaneous link to a corresponding electrode in the cochlea. Stimuli derived from channels with low center frequencies for the bandpass filter are directed to apical electrodes in the implant, and stimuli derived from channels with high center frequencies are directed to basal electrodes in the implant.

As with the IP, F0/F2, F0/F1/F2, and MPEAK processors, stimuli for the CIS strategy are delivered nonsimultaneously across electrodes. This eliminates one component of electrode or channel interactions that otherwise would be produced through direct summation of the electric fields from different (simultaneously stimulated) electrodes.

In contrast to the IP, F0/F2, F0/F1/F2, and MPEAK processors, the CIS strategy does not include an explicit representation of voiced or unvoiced distinctions or of the fundamental frequency of voiced sounds. Instead, voicing information is represented through relatively rapid variations in the modulation waveforms for each of the channels. These variations are included in the modulation waveforms with the use of high cutoff frequencies for the lowpass filters in the envelope detectors. A typical cutoff for CIS processors is in the range of 200 to 400 Hz, whereas a typical cutoff for the other processors is in the range of 20 to 35 Hz. The higher cutoff in CIS processors allows voicing information to pass through each envelope detector. The envelope signal in CIS processors conveys slow variations corresponding to changes in the shape of the vocal tract and the rapid variations corresponding to periodic, aperiodic, or mixed excitation of the tract. For voiced speech, the rapid periodic variations occur in synchrony with the periodic openings of the glottis.

CIS processors use higher rates of stimulation than the processors mentioned previously to represent adequately the rapid variations in the modulation waveforms. Rates for CIS processors typically exceed 500 pulses/s/electrode and often are much higher than that. Stimulus pulses for the other processors are delivered at the F0 rate during voiced speech and at a quasi-random or a somewhat higher fixed rate (*e.g.*, 300 pulses/s/electrode in variation 1 of the IP processor) during unvoiced speech.

The pulse rate for CIS processors must be higher than twice the cutoff frequency of the lowpass filters to avoid digital aliasing effects in the patterns of stimulation on single electrodes (Rabiner and Shafer, 1978; Wilson, 1997). Results from recent recordings of auditory nerve responses to sinusoidally amplitude modulated pulse trains indicate that the pulse rate should be even higher—four to five times the cutoff frequency—to avoid other distortions in the neural representations of modulation waveforms (Wilson, 1997 and 1999). A typical CIS processor might use a pulse rate of 1,000 pulses/s/electrode or higher, in conjunction with a 200-Hz cutoff for the lowpass filters.

The approach used in the CIS strategy departs from the vocoder-based approaches of prior strategies. No specific features of speech are extracted or represented with CIS processors. Instead, envelope variations in each of multiple bands are presented to the electrodes through modulated trains of interleaved pulses. The rate of stimulation for each channel and electrode does not vary between voiced and unvoiced sounds. This waveform or filterbank representation does not make any assumptions about how speech is produced or perceived. Instead, it seeks to represent the *acoustic environment* in a way that uses to the maximum extent possible the perceptual abilities of implant patients. The amount of information presented with CIS processors is much higher than the amounts presented with the prior approaches based on vocoder models. The information is tailored to fit within the perceptual spaces of electrically evoked hearing. The maximum frequency of envelope variations does not exceed the pitch saturation limit of typical patients (200 to 400 Hz), and the stimuli are interlaced to eliminate a principal component of electrode interactions. High pulse rates are used to represent without distortion the highest frequencies in the modulation waveforms. The modulation waveforms reflect rapid transitions in speech, including rapid consonant transitions and voicing information.

Stimuli for a simplified implementation of a four-channel CIS processor are shown in Fig. 7.10. The organization of this figure is the same as that of Fig. 7.5, which shows stimuli for variation 1 of the IP processors. The rate of stimulation is much higher for the CIS processor. The rate does not vary between voiced and unvoiced sounds, and the rate is adequate to represent most or all of the frequency information that can be perceived within channels.

An expanded display of stimuli during a 3.3-millisecond segment of the vowel input is presented in Fig. 7.11. This display shows the pattern of stimulation across electrodes. In this particular implementation of a CIS processor, stimulus pulses are delivered in a nonoverlapping sequence from the basal-most electrode (electrode 4) to the apical-most electrode (electrode 1). The rate of pulses on each electrode may be varied through manipulations in the duration of the pulses and time between sequential pulses. Any ordering of electrodes may be used in the stimulation sequence, such as an apex-to-base order or a staggered order (*i.e.*, an order designed to produce on average the maximum spatial separation between sequentially stimulated electrodes).

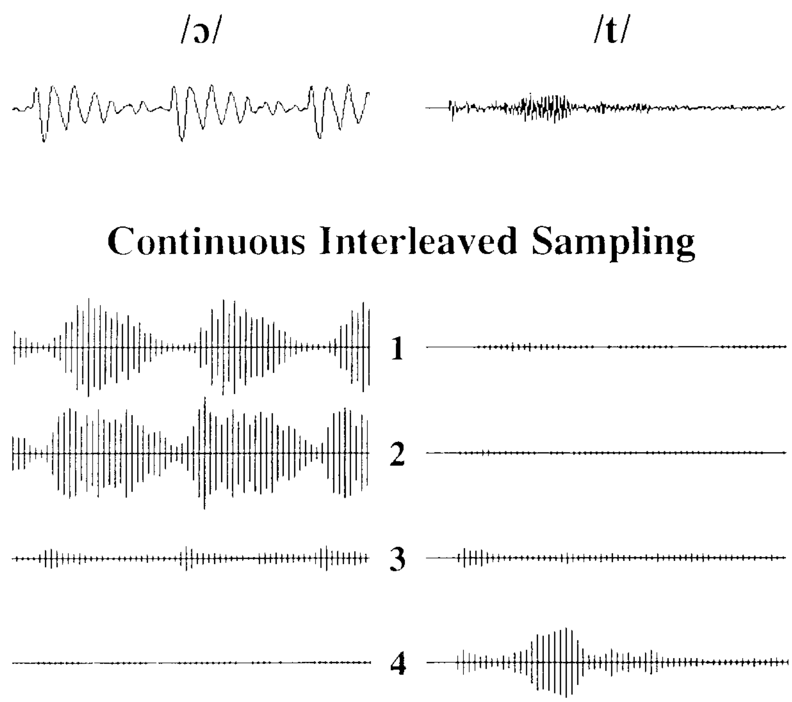

FIG. 7.10. Stimuli produced by a simplified implementation of a continuous interleaved sampling processor. The organization and speech inputs for this figure are the same as those in Fig. 7.5. (Adapted from Wilson *et al.*, 1991a, with permission.)

The first evaluations of the CIS strategy included comparisons among CA, IP (variation 1), and CIS processors for two subjects implanted with the Ineraid device. The percutaneous connector of the Ineraid device allowed high-quality implementations of each of the strategies. The results from tests of open-set recognition of words and sentences are presented in Fig. 7.12. Scores for the CIS processor were higher than the scores for the CA and IP processors for both subjects. For example, subject SR1 scored 8%,

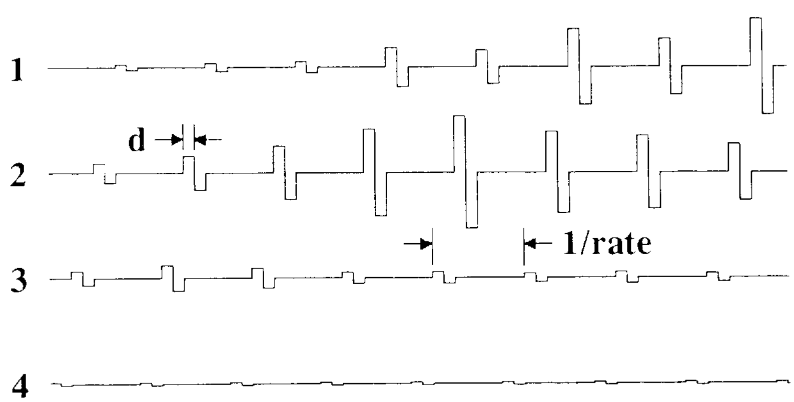

FIG. 7.11. Expanded display of continuous interleaved sampling stimuli. Pulse duration per phase (d) and the period between pulses on each electrode (1/rate) are indicated. The sequence of stimulated electrodes is 4-3-2-1. The duration of each trace is 3.3 ms. Notice that the pulses are nonsimultaneous, eliminating a principal component of electrode interactions due to summation of electric fields from different electrodes. (Adapted from Wilson *et al.*, 1991a, with permission.)

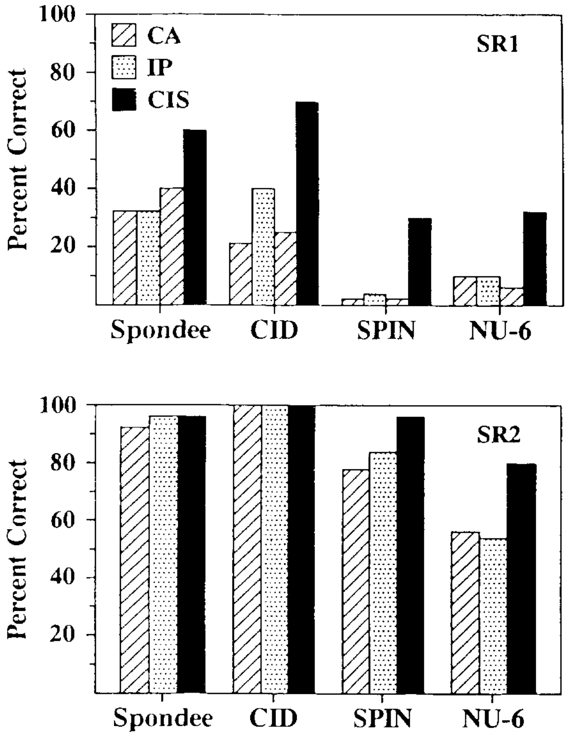

FIG. 7.12. Speech test scores for compressed analog (CA, *striped bars*), interleaved pulses (IP, *stippled bars*), and continuous interleaved sampling (CIS, *solid bars*) processors. Scores for Ineraid subject SR1 are presented in the top panel, and scores for Ineraid subject SR2 are in the bottom panel. The two CA scores for each test for subject SR1 are from separate evaluations of that processor. The first evaluation (4/89) was contemporaneous with the evaluation of the IP processor for this subject, and the second evaluation (8/90, *right-most CA bar for each test*) was contemporaneous with the evaluation of the CIS processor. Separate evaluations of the CA processor were not conducted for subject SR2, because evaluations of the IP and CIS processors were contemporaneous with the (single) evaluation of the CA processor. The tests included recognition of two-syllable words (Spondee), key words in the Central Institute for the Deaf (CID) sentences of everyday speech, the final word in each of 50 high-predictability sentences from the Speech Perception in Noise (SPIN) test (presented in these studies without noise), and monosyllabic words from Northwestern University Auditory Test 6 (NU-6). The CA processors for both subjects used four channels of processing and stimulation, and the IP and CIS processors for both subjects used six channels. The CA processors were used by the subjects in their daily lives. (From Wilson, 1993, with permission.)

10%, and 32% correct in recognizing NU-6 words with the CA, IP, and CIS processors, respectively, and subject SR2 scored 56%, 54%, and 80% correct in recognizing those words for the same processors (Lawson *et al.*, 1993; Wilson, 1993). The subjects had used the CA processor for two or more years in their daily lives at the time of these tests and had had some limited (laboratory) experience with the IP processor. Experience with the CIS processor was no more than several hours for each of the subjects. Despite this relative lack of experience, both subjects achieved their highest scores with CIS for all tests that were sufficiently difficult not to produce scores at or near 100% correct for all three processors. The score of 80% correct for recognition of the NU-6 words was unprecedented at the time of this study. The immediate jumps in scores with these initial implementations of the CIS strategy encouraged its further development and subsequent tests with additional subjects.

Compressed Analog Strategy

The CA strategy was among the first strategies used with multi-electrode implants (Eddington, 1980; Merzenich *et al.*, 1984). Its development was roughly contemporaneous with the development of the F0/F2 strategy and predated the development of the IP strategies.

The CA strategy has been applied in conjunction with the now-discontinued Ineraid and UCSF/Storz implants. A modification of the strategy, *simultaneous analog stimulation* (SAS), is used as one of several strategies that can be selected for the Clarion (Advanced Bionics) implant. The differences between CA and SAS are described in a separate section on SAS.

In contrast to the other strategies described in this chapter, the CA and SAS strategies use analog waveforms for stimuli instead of biphasic pulses. A block diagram of the CA strategy is presented in Fig. 7.13. A microphone or other input is compressed with a fast-acting AGC. The AGC output is

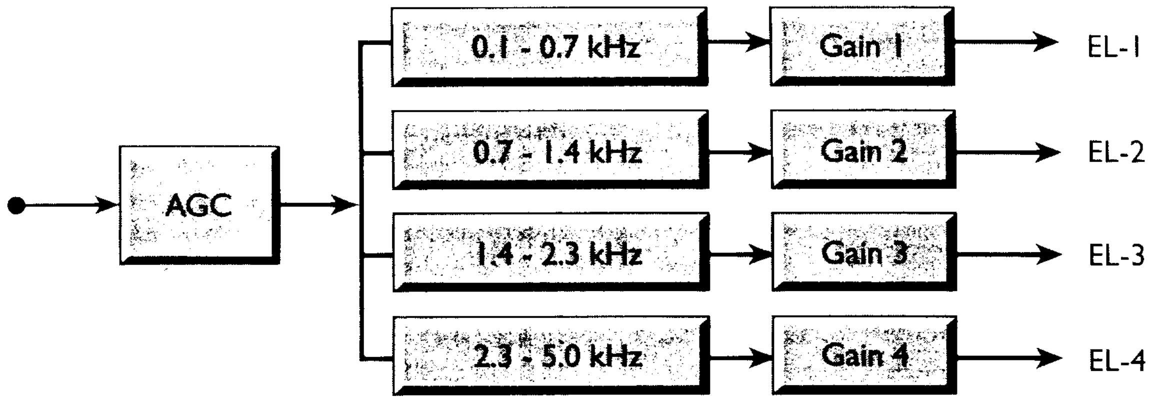

FIG. 7.13. Diagram of the compressed analog processing strategy. The CA strategy uses a broadband automatic gain control (AGC), followed by a bank of bandpass filters. The outputs of the filters can be adjusted with independent gain controls. Compression is achieved through rapid action of the AGC, and high-frequency emphasis and limited mapping to individual electrodes is accomplished through adjustments of the channel gains. (Adapted from Wilson *et al.*, 1991a, with permission.)

filtered into contiguous bands (usually four) that span the range of speech frequencies. The signal from each bandpass filter is amplified and then directed to a corresponding electrode in the implant. The gains of the amplifiers for the different bandpass filters can be adjusted to produce stimuli that do not exceed the upper end of the dynamic range of percepts for each electrode and that provide a high-frequency emphasis (*e.g.*, percepts for high-frequency channel 4 can be made as loud as percepts for low-frequency channel 1 even though high-frequency sounds in speech generally are much less intense than low-frequency sounds).

The compression provided by the AGC reduces the dynamic range of the input to approximate the dynamic range of electrically evoked hearing. The dynamic range of stimulation also can be restricted with a clipping circuit before or after the amplifiers for some or all of the bandpass channels (Merzenich, 1985; Merzenich *et al.*, 1984).

Stimuli produced by a simplified implementation of a CA processor are shown in Fig. 7.14. The format of this figure is the same as that in Figs. 7.5 and 7.10, which show stimuli for IP and CIS processors, respectively. Comparisons of the figures show a highly detailed representation for the CA processor, a somewhat less detailed representation for the CIS processor, and a relatively sparse representation for the IP processor. The CA and CIS processors do not extract nor explicitly represent features of speech. The IP processor extracts voiced or unvoiced boundaries and the fundamental frequency of voiced speech sounds. These features are represented by rates of stimulation, which are different between the panels for voiced and unvoiced inputs in Fig. 7.5.

CA stimuli represent a large portion of the information in unprocessed speech input. Spectral and temporal patterns of speech are represented in the relative amplitudes of the stimuli across electrodes and in the temporal variations of the stimuli for each of the electrodes. The approach is to present as much information as possible to the implant and to allow the brain to extract what it can from this minimally processed representation.

Like the CIS strategy, the CA strategy is not based on a vocoder model. Neither strategy extracts features from speech inputs. Both strategies present a range of temporal variations in each channel (and at each electrode) that at least includes the range of voicing information in speech. The ranges of temporal variations within channels for CA and CIS processors are much greater than the range within channels for a typical channel vocoder.

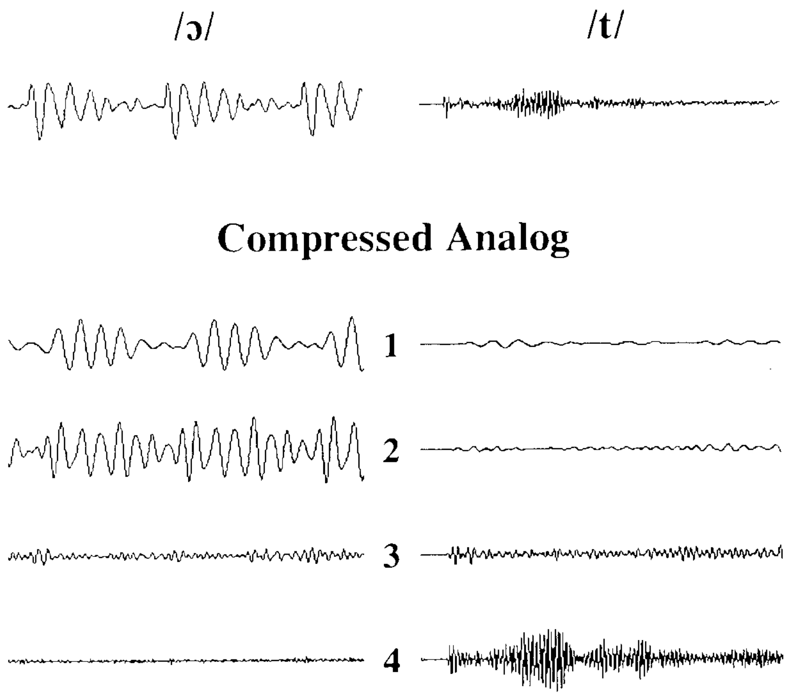

FIG. 7.14. Stimuli produced by a simplified implementation of a compressed analog processor. The organization and speech inputs for this figure are the same as those in Figs. 7.5 and 7.10. (Adapted from Wilson *et al.*, 1991a, with permission.)

Such departures from a vocoder model recognize that the principal goal of vocoder designs is different from the principal goal of implant designs. The goal for vocoders is to minimize the information rate of transmission while still supporting an acceptable intelligibility of the reconstructed speech signal at the receiver. The goal for implants is to maximize the amount of speech information that can be perceived and used by the recipient. This fundamental difference can lead to different choices in design.

Until the advent of the CIS strategy, performance with the CA strategy was at least as good as the performance with any other processing strategy for commercially available multi-electrode implants (Cohen *et al.*, 1993; Dorman, 1993; Dorman *et al.*, 1989; Gantz *et al.*, 1987 and 1988; Tye-Murray and Tyler, 1989; Tyler *et al.*, 1995; Tyler and Moore, 1992). It supported some open-set recognition of speech for a relatively large fraction of patients. In one group of 50 Ineraid patients, for example, recognition of NU-6 words ranged from 0% to 60% correct, with a median score of 14% correct (Dorman *et al.*, 1989). Comparisons of scores for different groups of subjects using the Nucleus implant (with the F0/F2 or F0/F1/F2 processing strategies) or the Ineraid implant (with the CA processing strategy) showed no difference for a wide variety of speech reception tests in quiet. In some studies, scores for tests in noise were significantly higher for the CA strategy (Gantz *et al.*, 1987; Tyler *et al.*, 1995). The relative advantage of the CA strategy in noise may have been a result of the high sensitivity of the feature extraction circuits in the Nucleus processors to noise interference.

Although performance with the CA strategy was among the best before the introduction of CIS, substantial open-set recognition by most patients still had not been achieved. The average user struggled to understand even high-context sentences with hearing alone, and very few patients were able to use the telephone except in constrained situations involving exchanges of previously arranged codes.

Comparisons of Compressed Analog and Continuous Interleaved Sampling Strategies

The CIS strategy was designed to eliminate or at least ameliorate some possible weaknesses in the CA strategy. A major concern with the CA approach is that simultaneous stimulation of multiple electrodes can produce large and uncontrolled interactions through vector summation of the electric fields from each of the electrodes (White *et al.*, 1984). The resulting degradation of independence among electrodes and channels would be expected to reduce the salience of channel-related cues. In particular, the neural response to stimuli presented at one electrode may be significantly distorted or even counteracted by coincident stimuli presented at other electrodes. The pattern of interaction also may vary according to the instantaneous phase relationships among the stimuli for each of the electrodes. Phase is not controlled within or across channels in CA processors, and this may degrade further the representation of the speech spectrum according to place of stimulation.

The problem of electrode interactions is addressed in CIS processors with the use of nonsimultaneous stimuli. This eliminates a principal component of interactions due to of direct summation of electric fields from different electrodes. It does not eliminate another component, often called *temporal channel interactions.* These interactions refer to effects of preceding stimuli on the response to a given stimulus. They may be produced by the refractory properties of neurons or by temporal summation of sequential stimuli at neural membranes. In the first case, a stimulus at one electrode can excite a population of neurons that overlaps the population that would be excited by another electrode, if not preceded by the stimulus at the first electrode. The neurons excited by the first stimulus may still be in a refractory state at the time of the second stimulus. Neurons in their absolutely refractory period cannot be excited by a second stimulus, and neurons in their relatively refractory period may or may not be excited by a second stimulus. Neurons in their relatively refractory period are less excitable (*i.e.*, have higher thresholds) than neurons in a resting state. Stimulation of neurons by a prior stimulus or prior stimuli can alter the number and distribution of neurons stimulated by a following stimulus when the following stimulus falls within the refractory periods of the previously stimulated neurons (typical auditory neurons have absolute refractory periods of about 0.5 milliseconds and relatively refractory periods with time constants in the range of 0.5 to 4.0 milliseconds).

The number and distribution of neurons excited by a stimulus pulse also can be altered by preceding stimuli through temporal summation effects. In this mechanism, a subthreshold depolarization of a neural membrane can be augmented with a subsequent depolarization if the second stimulus is presented within the temporal integration window of the membrane. The integration window is determined by the time constant of the membrane, which is on the order of 50 to 200 microsecond for myelinated mammalian neurons. An accumulation of subthreshold depolarizations ultimately may excite the neuron that would not have been excited with the final stimulus alone. Refractory effects can reduce the number of neurons excited with a final stimulus compared with the number excited with that stimulus in the absence of preceding stimuli, whereas summation effects can increase the number. Either change distorts the intended pattern of stimulation for a given electrode.

Fortunately, temporal channel interactions appear to be small compared with the interactions produced with simultaneous stimulation of electrodes (Favre and Pelizzone, 1993). The relatively small effects of temporal channel interactions may be reduced further through a staggered order of stimulation across electrodes in CIS and other pulsatile processors. Such orders are designed to produce the maximum possible spatial separation (on average) between sequentially stimulated electrodes (*e.g.*, an order of 6-3-5-2-4-1 for a six-channel CIS processor). In general, a greater spatial separation should reduce the overlaps in the excitation fields between electrodes and thereby reduce the magnitudes of the interactions.

An additional possible weakness of CA processors is that only a small part of the presented information can be perceived. For example, patients cannot perceive differences in frequencies of electrical stimulation on single electrodes above the pitch saturation limit, usually between 200 and 400 Hz. Thus, many of the temporal details in CA stimuli are not likely to be accessible to the typical user. Presentation of information outside the perceptual space does not help the patient and may be destructive if it gets in the way of something else or if it distorts or interferes with the information that can be perceived. Presentation of highly detailed temporal information with analog stimuli, for instance, precludes a reduction in electrode interactions that might otherwise be obtained with nonsimultaneous pulses.

The cutoff frequency of the lowpass filters in the envelope detectors of CIS processors is set to include most or all of the frequencies that can be perceived as different pitches by implant patients (typical cutoffs of 200 to 400 Hz). Rates of stimulation at each electrode are at least high enough to prevent aliasing (twice the cutoff frequency of the lowpass filters) and usually much higher than that to eliminate other distortions in the neural representation of modulation waveforms.

A further concern with the CA strategy is its front-end compression. The signal presented to the bank of bandpass filters is a highly compressed version of the signal at the input to the processor. A typical compression ratio is 6 to 1 (Merzenich *et al.*, 1984). Such high levels of compression produce harmonics and other distortions in the spectrum of the input signal, sometimes called *spectral splatter*. These distortions are transmitted through the bank of bandpass filters and ultimately to the electrodes (White, 1986). They most likely degrade the representation of across-channel cues to the speech spectrum. For example, spurious components at high frequencies are produced for all speech sounds that have significant energy at low frequencies.

CIS processors do not use an AGC or use one with relatively slow attack and release times and with a low compression ratio. This eliminates or greatly reduces the spectral distortions described above.

Compression in CIS processors is accomplished with the logarithmic or power-law mapping functions for each of the electrodes. This back end compression eliminates the spectral distortions produced with front end compression and allows a precise matching of the full range of envelope signals in each channel to the dynamic range (from threshold to a comfortably loud or loud level) of the corresponding electrode. The logarithmic function produces a normal or near-normal growth of loudness with increases in sound level (Dorman *et al.*, 1993; Eddington *et al.*, 1978; Zeng and Shannon, 1992). The channel-by-channel mapping of acoustical dynamic range onto the electrical dynamic range uses fully the discriminable steps in amplitude available for each electrode.

The CA and CIS strategies have been compared in a number of studies, beginning with a study conducted by Wilson *et al.* in 1989 and 1990 (Wilson *et al.*, 1991a). In that study, a laboratory implementation of the CIS strategy was compared with the clinical CA processor of the Ineraid device. Eleven subjects participated in the study. Seven of the subjects were selected for their high levels of performance with the CA processor and Ineraid implant, and four additional subjects were subsequently selected for their low levels of performance with that processor and implant. The high-performance subjects were representative of the best results in terms of speech reception scores that had been obtained with the Ineraid or any other implant system at the time of the study. All subjects had had at least 1 year of daily experience with the clinical CA processor when the comparisons were conducted, and most had had multiple years of such experience. In contrast, experience with CIS was limited to several days of initial tests to evaluate fitting alternatives. Experience with the CIS processor ultimately compared with the clinical CA processor was no more than several hours.

Results from the study are presented in Fig. 7.15. Lines connect CA scores with CIS scores for each subject. The four panels show results for different open-set tests. Scores for the high-performance subjects are indicated by the endpoints of the light lines in the upper part of each panel, and scores for the low-performance subjects are indicated by the end points of the thicker lines closer to the bottom of each panel. All tests were conducted with hearing alone, and all tests used single presentations of recorded material without feedback about correct or incorrect responses.

Scores for all 11 subjects were improved with the CIS processor. The average scores across subjects increased from 57% to 80% correct in the spondee tests (*i.e.*, recognition of two-syllable words), from 62% to 84% correct in the CID sentence tests, from 34% to 65% correct in the Speech Perception in

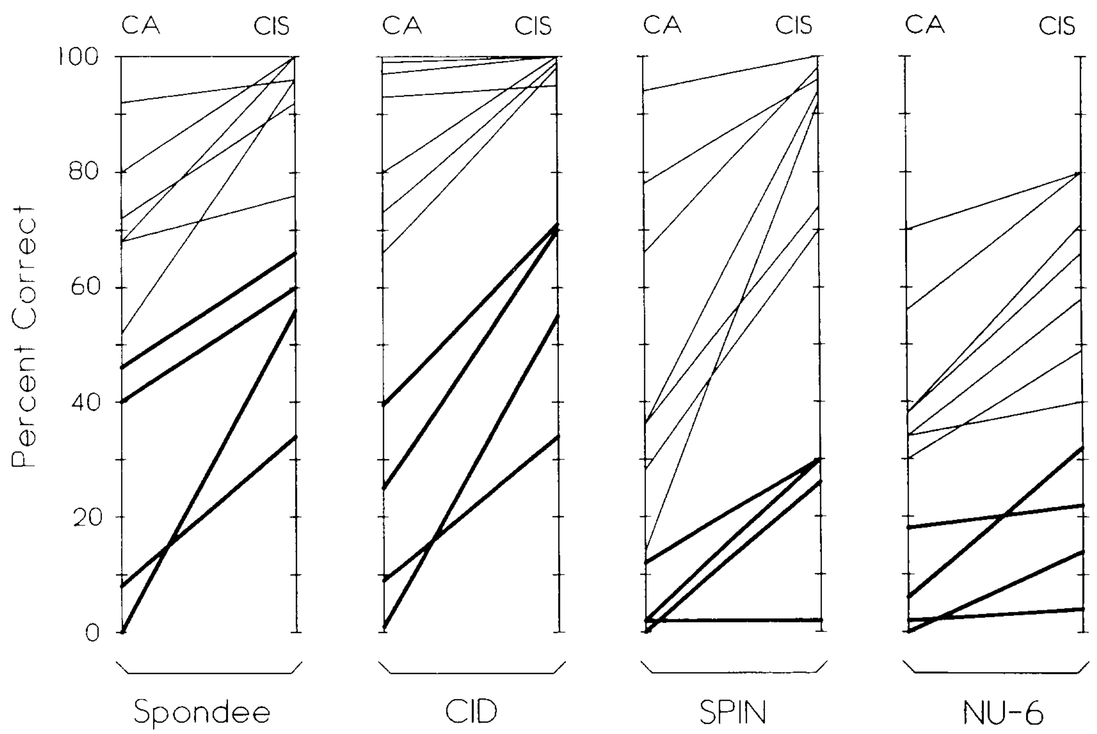

FIG. 7.15. Speech recognition scores for compressed analog (CA) and continuous interleaved sampling (CIS) processors. A line connects the CA and CIS scores for each subject. Light lines correspond to seven subjects selected for their excellent performance with the clinical CA processor of the Ineraid device, and the heavier lines correspond to four subjects selected for relatively poor performance. The CA strategy provided performance that was equal to or better than all clinically available alternatives at the time of these comparisons, which began in 1989. The tests included recognition of two-syllable words (Spondee), key words in the Central Institute for the Deaf (CID) sentences of everyday speech, the final word in each of 50 high-predictability sentences from the Speech Perception in Noise (SPIN) test (presented in these studies without noise), and monosyllabic words from Northwestern University Auditory Test 6 (NU-6). (Adapted from Wilson *et al.*, 1995, with permission.)

Noise (SPIN) sentence tests, and from 30% to 47% correct in the NU-6 word tests. All of these increases were highly significant.

The scores of 80% correct on the NU-6 test for two of the subjects were unprecedented at the time of this study. Perhaps even more important were the improvements produced with CIS for the low-performance subjects. Subject SR1, for instance, achieved scores with the CIS processor that moved him from a low-performance to a high-performance category. His scores improved from 40% to 60% correct for the spondee tests, from 25% to 70% correct for the CID sentence tests, from 2% to 30% correct for the SPIN sentence tests, and from 6% to 32% correct for the NU-6 tests. Similarly, results for subject SR10 demonstrated quite large improvements with CIS. His scores improved from 0% to 56% correct for the spondee tests, from 1% to 55% correct for the CID sentence tests, from 0% to 26% correct for the SPIN sentence tests, and from 0% to 14% correct for the NU-6 tests. This subject went from no open-set recognition of speech to useful open-set recognition with substitution of a CIS processor for the clinical CA processor.

Results from subsequent studies (*e.g.*, Böex *et al.*, 1994 and 1996; Dorman and Loizou, 1997a and 1997b) have replicated and extended these initial findings. Subsequent studies have demonstrated a potential for additional increases in performance with adjustments in the parameters of CIS processors (Wilson *et al.*, 1995) and with further experience in using CIS processors (Dorman and Loizou, 1997a; Lawson *et al.*, 1995; Pelizzone *et al.*, 1995). One of the subjects who scored 80% correct on the NU-6 test using the CIS processor of the initial comparisons scored in the high 90th percentile with his clinical CIS processor. Subject SR10 scores in the 60th percentile with his refined fitting of a CIS processor and with the experience he has gained using the processor in his daily life.

The immediate improvements observed with substitution of a laboratory CIS processor for the clinical CA processor may have been produced by one or more of the differences in design described previously. In addition, five or six channels (and electrodes) were used for the CIS processors, whereas the standard four were used with the clinical CA processor. This increase also might have contributed to the better performance of the CIS processors.

Spectral Maxima Sound Processor Strategy

The progenitor of the SPEAK strategy was the *spectral maxima sound processor* (SMSP) strategy. It and SPEAK mark a departure from the feature extraction approach of prior strategies used in conjunction with the Nucleus implant.

The SMSP strategy is an *n*-of-*m* strategy, as described in the section on IP processors. The patent for the SMSP builds on the variation 2 of IP processors by specifying particular choices for *n* and *m* and by specifying a higher cutoff frequency for the lowpass filters in the envelope detectors (McDermott and Vandali, 1997).

In the SMSP (McDermott *et al.*, 1992; McDermott and Vandali, 1997), the microphone or direct input is directed to an AGC (with long time constants and a low compression ratio), and the AGC output is directed to a bank of 16 bandpass filters spanning the range from 250 to 5,400 Hz. Envelope signals are derived for each of the bandpass outputs with a rectifier and lowpass filter, as in the IP, F0/F2, F0/F1/F2, and MPEAK strategies. In contrast to those strategies, the cutoff frequency for the lowpass filters is set at 200 Hz instead of a much lower value. As with the IP strategies, each bandpass channel is assigned to a position along the electrode array. Channels with low center frequencies deliver their outputs to electrodes at apical positions, and channels with high center frequencies deliver their outputs to electrodes at basal positions. A postprocessor is used to scan the outputs of the bandpass channels (*i.e.*, the envelope signals) for each cycle of stimulation across electrodes. The scan first selects channel outputs above a preset noise threshold, and if more than six channel outputs are selected, it selects the six channel outputs that have greatest amplitudes among the set. The identified channels and associated electrodes (usually six except for quiet conditions or weak sounds of speech) are used for stimulation in that cycle. A logarithmic transformation of the envelope signals is used to determine pulse amplitudes, as in the prior strategies described earlier. The electrodes are stimulated in a nonsimultaneous sequence for each cycle, starting with the electrode or electrode pair assigned to the channel with the highest envelope signal among the selected channals, and ending with the electrode or electrode pair assigned to the channel with the lowest envelope signal among the selected channels. Stimulation cycles are repeated at the rate of 250/s, which approximates the maximum rate possible for stimulation of six electrodes or six electrode pairs in each cycle with the Nucleus transcutaneous link. (A rate of around 400/s can be attained under certain conditions, see Crosby *et al.*, 1985, and Shannon *et al.*, 1990).

As with the CIS strategy, voicing information may be represented with the SMSP strategy through variations in modulation waveforms out to 200 Hz. However, the maximum pulse rate on each electrode with the SMSP strategy is only 250/s, which is below the aliasing rate of 400 pulses/s for a 200-Hz variation in the modulation waveform. The representation of frequencies above 125 Hz is subject to aliasing effects with the SMSP, and the representation of frequencies in the range of one-fourth to one-half the pulse rate probably is distorted to a lessor extent (Busby *et al.*, 1993; McKay *et al.*, 1994; Wilson, 1997). Such aliasing effects and distortions are likely to degrade the representation of voicing information and also produce reversals in judgments of fundamental frequencies when the frequencies exceed one-half the carrier rate (*i.e.*, 125 Hz in the middle part of the F0 range for male talkers and in lower part of the F0 ranges for female and child talkers).

Results from tests with a small number of subjects indicated that the SMSP strategy might offer substantial improvements in speech reception compared with the F0/F1/F2 and MPEAK strategies (McDermott *et al.*, 1992; McKay *et al.*, 1991 and 1992). For example, in studies with four subjects, McKay *et al.* (1992) recorded significant increases in several measures of speech reception when the SMSP was substituted for the MPEAK processor. The measures included identification of vowels and consonants, recognition of consonant-vowel nucleus-consonant (CNC) monosyllabic words, and recognition of BKB sentences in competition with multitalker babble at the speech-to-babble ratio of 10 dB. The average scores across subjects improved from 76.3% to 91.3% correct for vowels, from 59.4% to 74.9% correct for consonants, from 39.9% to 57.4% correct for words, and from 50.0% to 78.7% correct for sentences in noise.

Spectral Peak Strategy

After the encouraging results with the SMSP, Cochlear Ltd. and the University of Melbourne developed the SPEAK strategy, which is a refinement of the SMSP strategy. New hardware also was developed to implement the SPEAK strategy in a smaller package and with more processing options than the MSP (Patrick *et al.*, 1997).

In the SPEAK strategy (Patrick *et al.*, 1997; Skinner *et al.*, 1994), the input is filtered into as many as 20 bands rather than the 16 bands of the SMSP. Envelope signals are derived with a rectifier and lowpass filter, as in the SMSP and prior strategies. The cutoff of the lowpass filters is set at 200 Hz, as in the SMSP. The outputs of the envelope detectors are scanned with a postprocessor in a manner similar to that of the SMSP. The number of bandpass channels selected in each scan depends on the number of envelope signals exceeding a preset noise threshold and on details of the input such as the distribution of energy across frequencies. In many cases, six channels are selected, as in the SMSP strategy. However, the number can range from one to a maximum that can be set as high as

10. Cycles of stimulation, which include the selected channels and associated electrodes, are presented at rates between 180 and 300/s. The amount of time required to complete each cycle depends on the number of electrodes and channels included in the cycle and the pulse amplitudes and durations for each of the electrodes. In general, inclusion of relatively few electrodes in a cycle allows relatively high rates, whereas inclusion of many electrodes reduces the rate.

A diagram illustrating the operation of the SPEAK processor is presented in Fig. 7.16. The speech input is directed to a bank of bandpass filters and envelope detectors, whose outputs are scanned for each cycle of stimulation. In this diagram, six channels are selected in the scan and the corresponding electrodes are stimulated nonsimultaneously in a base-to-apex order. Two such scans are depicted in the figure.

The new hardware, called the Spectra 22 processor, includes a custom integrated circuit to perform the functions of bandpass filtering and envelope detection. The integrated circuit can be programmed to produce changes in the frequency ranges and gains for the bandpass filters. In a typical implementation of the SPEAK strategy, 20 bandpass filters span the range from 150 to 10,823 Hz, with a linear spacing of bandpass frequencies below 1,850 Hz and a logarithmic spacing of bandpass frequencies above 1,850 Hz. (Such linear-logarithmic spacing may provide an even closer match to the distribution of critical bands in normal hearing compared with a strictly logarithmic spacing.) The filter gains normally are all set to a single value. Alternative choices may be specified, such as when fewer than 20 electrode positions are available for a given patient.

Comparisons between Spectral Peak and Multipeak Strategies

Within-subject comparisons of the SPEAK and MPEAK strategies have been conducted with 63 English-speaking patients at various centers in Australia, the United States, Canada, and England (Skinner *et al.*, 1994). Some of the principal results and key processing steps in the SPEAK strategy are presented in Fig. 7.17. Subjects were tested using an ABAB crossover design, in which subjects used their clinical MPEAK processor during an initial 3-week period, then used the SPEAK processor for 6 weeks, then returned to the MPEAK processor for 3 weeks, and then used the SPEAK processor during a final 3-week period. The processors were tested at the end of each period. All subjects had had at least 8 months of daily experience with the MPEAK processor before the study.

The subjects participating in the study all had scores of 5% correct or better for recognition of key words in the CID or BKB sentences. This group of subjects represented approximately 75% of the population using the Nucleus device at that time. (The remaining 25% of the population had little or no open-set recognition of speech, even for relatively easy tests such as the sentence tests described earlier.)

Averages of the scores for each subject and processor are presented in the middle and bottom panels of Fig. 7.17. The middle panel shows scores for the recognition of key words in the City University of New York (CUNY) sentences or in the Speech Intelligibility Test for Deaf Children (SIT) sentences presented in quiet. The bottom panel shows scores for the sentences presented in competition with multitalker speech babble at the speech-to-babble ratio of 10 dB. Two of the 63 subjects were excluded from the tests of sentence recognition in quiet, and 5 were excluded from the tests of sentence recognition in noise because of their low scores in preliminary tests using the CID sentences; the excluded subjects scored below 35% correct on the CID test.

Among the 61 subjects who were tested with sentences in quiet, 24 had significantly higher scores with the SPEAK processor, and 2 had significantly higher scores with the MPEAK processor. Many of the subjects obtained high scores on this relatively easy test and therefore possible differences be-

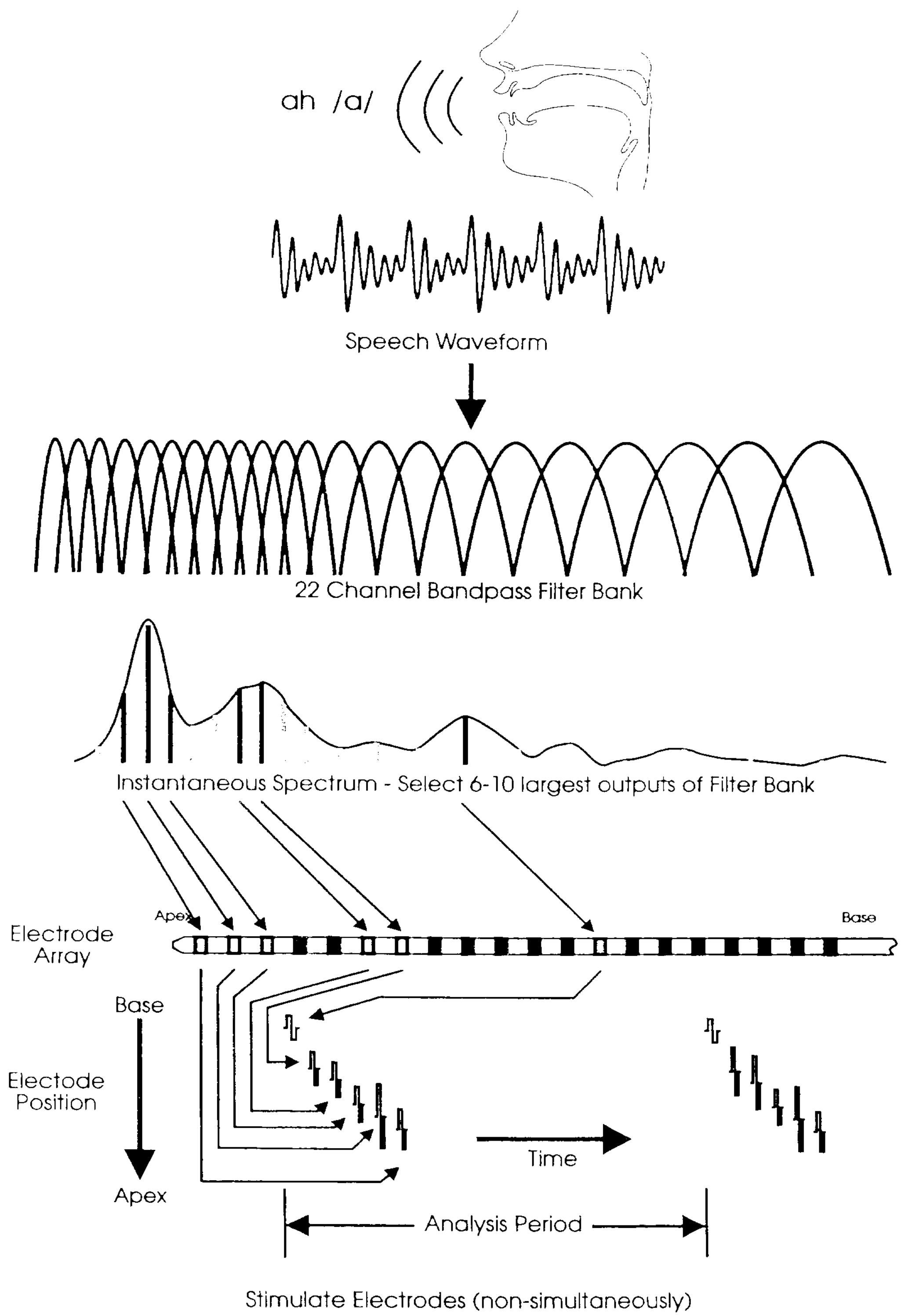

FIG. 7.16. Key steps in the spectral peak processing strategy. Speech inputs are directed to a bank of up to 20 bandpass filters and envelope detectors. A postprocessor scans the envelope signals for each cycle of stimulation across electrodes. Between 1 and 10 of the highest-amplitude signals are selected in each scan, depending on characteristics of the input (*i.e.*, overall level and spectral composition). Electrodes associated with the selected envelope signals and bandpass channels are stimulated in a base-to-apex order. (From Patrick *et al.*, 1997, with permission.)

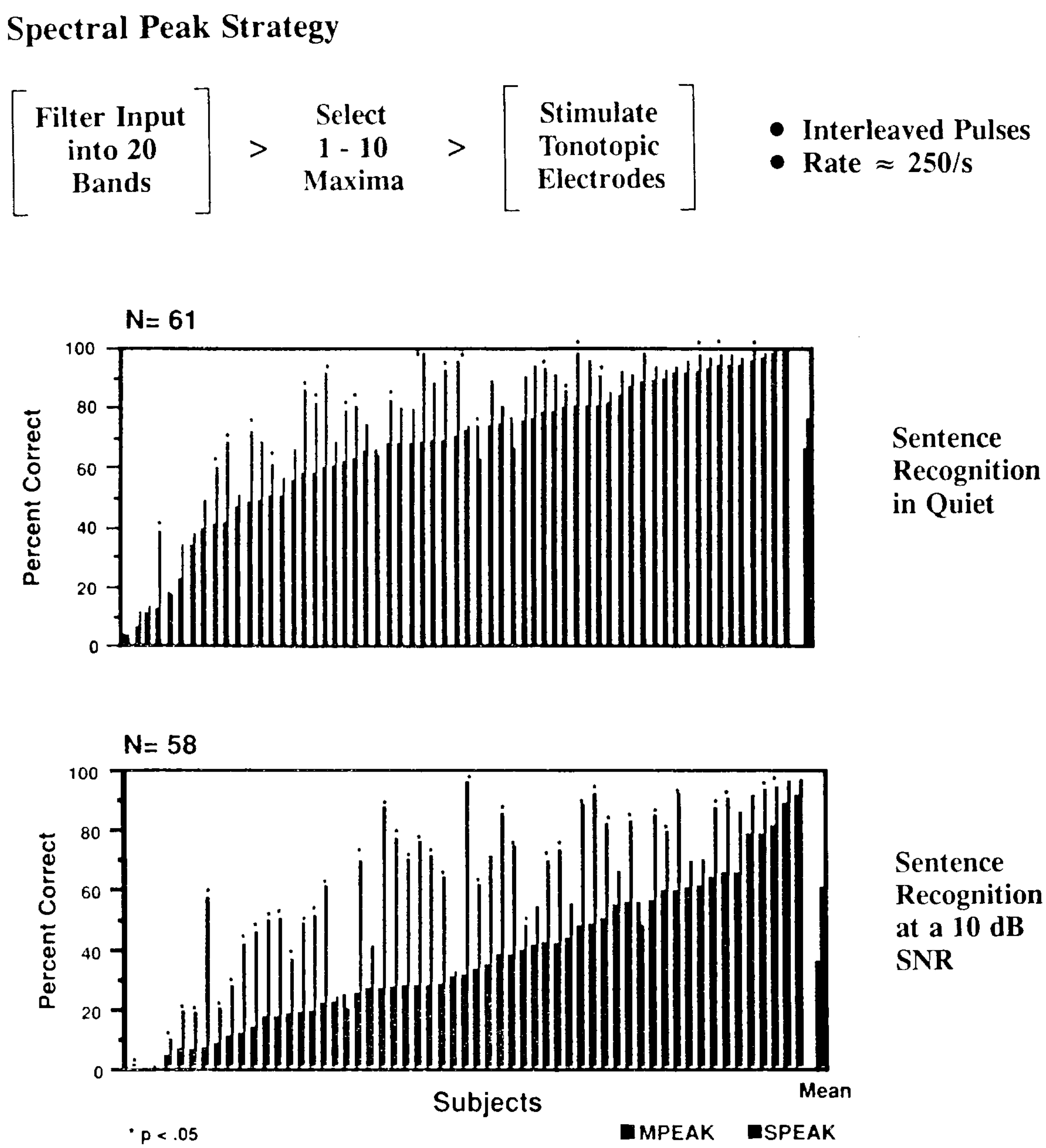

FIG. 7.17. Mean scores for recognition of key words in City University of New York sentences or in the Speech Intelligibility Test for Deaf Children sentences with multipeak (MPEAK) and spectral peak (SPEAK) processors. The top panel indicates the design of the SPEAK strategy. The MPEAK strategy was used in the clinical Nucleus processor at the time of these comparisons. The MPEAK strategy operated in a fundamentally different way from SPEAK, extracting and representing specific features of the speech input, such as formant frequencies and amplitudes. The middle panel shows scores for each of 61 subjects listening to the sentences in a condition without interfering noise, and the bottom panel shows scores for each of 58 subjects listening to the sentences in competition with multitalker-babble noise at the speech-to-babble ratio of 10 dB. (Adapted from Wilson *et al.*, 1995, with permission; data from Skinner *et al.*, 1994)

tween processors may have been masked by ceiling effects. Among the 58 subjects who were tested with sentences in noise (for the 10 dB speech-to-babble condition), 41 had significantly higher scores with the SPEAK processor, and none had a significantly higher score with the MPEAK processor.

The principal advantage of the SPEAK processor, as demonstrated by these and other tests (Skinner *et al.*, 1994), is in better recognition of speech in noise. This advantage may be a result of the filterbank approach used in the SPEAK processor. It also may be a result of the high sensitivity of the feature extraction portions of the MPEAK processor to noise. The accuracy of such extraction can be severely degraded by even small amounts of noise. Zero crossings analysis in particular is susceptible to deleterious effects of noise interference (Rabiner and Shafer, 1978).

Comparisons between Spectral Peak and Continuous Interleaved Sampling Strategies

Comparisons of SPEAK and CIS strategies indicate that both support high levels of speech reception in quiet conditions (Aubert *et al.*, 1998; Kiefer *et al.*, 1996; Lawson *et al.*, 1996; Loizou *et al.*, 1997). Comparisons with matched groups of patients in the study of Kiefer *et al.* (1996) produced significantly higher scores with CIS (as implemented in the Med El COMBI 40 device) for recognition of key words in the Göttingen sentences at the two tested speech-to-noise ratios of 15 and 10 dB. Scores for recognition of key words in the easier Innsbruck sentences were not significantly higher with CIS for presentation of the sentences in quiet and at the tested speech-to-noise ratios of 15 and 10 dB. Scores for the Innsbruck sentences were high for the CIS strategy for the three test conditions, and therefore the comparisons between the two strategies with those sentences may have been limited by possible ceiling effects. The average scores for the CIS strategy were 88%, 81%, and 72% correct for the quiet, 15 dB and 10 dB conditions, respectively, and the average scores for the SPEAK strategy were 79%, 66%, and 56% correct for those same conditions.

Comparisons with separate matched groups of patients in the study of Loizou *et al.* (1997) also indicated an advantage of CIS (as implemented in the CIS-Link processor developed by Med El for use with the Ineraid implant) for speech reception in noise. The tests included identification of 16 consonants in an /a/-consonant-/a/ context, identification of 8 vowels in an /h/-vowel-/d/ context, and recognition of key words in sentences from the Hearing in Noise Test (HINT) database. Each of the tests was administered without noise and at the speech-to-noise ratios of 15, 10 and 5 dB. The scores from the tests of consonant identification were significantly higher with CIS for all conditions. The score from the test of vowel identification for the most adverse speech-to-noise ratio (5 dB) also was significantly higher with CIS. Scores from the remaining tests were not significantly higher with CIS. Loizou *et al.* suggested that the results from the sentence tests were consistent with those reported by Kiefer *et al.*, in that the HINT sentences are comparable in difficulty to the Innsbruck sentences. Tests with more difficult material may have provided a greater sensitivity for demonstrating possible differences between the two strategies.

Lawson *et al.* (1996) evaluated CIS, *n*-of-*m*, and SPEAK processors in within-subject comparisons with five subjects. The tests included identification of 16 or 24 consonants in an /a/-consonant-/a/ context, presented without noise; the 24 consonant test was used for the two subjects with the highest levels of performance among processors to avoid possible ceiling effects. The results indicated an equivalence or superiority of six-channel CIS processors (as implemented with a laboratory system) compared with the clinical SPEAK processor used by the subjects in their daily lives. The CIS and SPEAK strategies produced scores that were not statistically different when the range of frequencies spanned by the bandpass filters in the CIS processors was from 350 to 5,500 Hz. The CIS strategy produced significantly higher scores when the upper end of the range was extended to 9,500 Hz, approximating the upper end of the range in the SPEAK processors.

In a separate study with 41 subjects, Aubert *et al.* compared the clinical SPEAK processor used by the subjects with a CIS processor, as implemented in the new Sprint hardware of the Cochlear Ltd. CI24M implant system. Some subjects obtained higher speech reception scores with the SPEAK strategy, and other subjects obtained higher scores with the CIS strategy. The remaining subjects had similar scores with the two strategies.

Comparisons of the CIS and SPEAK strategies have indicated an approximate equivalence of the two for some tests, such as the Innsbruck and HINT sentences, and a superiority of CIS for other tests, such as identification of consonants. CIS appears to provide an advantage for listening to speech in noise, particularly for consonants and difficult sentences.

N-of-*M* Strategies

Variation 2 of the IP strategy was the first *n*-of-*m* processor for cochlear implants. Since that beginning, the SMSP and SPEAK strategies have been developed for use with the Cochlear Ltd. implant, and an *n*-of-*m* strategy using much higher rates of stimulation than the SMSP or SPEAK strategies has been developed for use with the Med El COMBI 40 and COMBI 40+ implants. An *n*-of-*m* strategy using similarly high rates of stimulation has been developed as a processing option for use with the new Cochlear Ltd. CI24M device. This last implementation is called the ACE strategy.

The *n*-of-*m* approach may help in making more prominent the highest amplitude envelope signals among many. Results from early vocoder studies (Flanagan, 1972) indicated that envelope signals 30 dB below the peak signal can be discarded without damaging the intelligibility of speech transmitted and reconstructed with a channel vocoder and that a modest further reduction in the number of selected envelope signals still can support high levels of intelligibility while allowing a reduction in the information rate of transmission. The latter observation led to the development of peak-picker vocoders that applied various rules to select the peaks from among the envelope signals that would maximize intelligibility for a given number of peaks or minimize the transmission rate for a given target intelligibility. In general, the most effective rules did not involve simple selection of the *n* greatest envelope signals, but instead selected signals that would convey the essential spectral information without redundancy. (The performance of *n*-of-*m* strategies for cochlear implants may be improved with the use of such rules, but this possibility has not yet been evaluated.)

Lawson *et al.* (1996) compared processors implementing an *n*-of-*m* strategy along with other strategies in tests with five subjects implanted with the Nucleus electrode array and with percutaneous access to that array. The percutaneous access allowed implementations of strategies that combined relatively high rates of stimulation with a relatively large number of addressed electrodes. The *n*-of-*m* processors selected the six highest envelope signals from among 18. In one implementation, the cycles of stimulation across electrodes were repeated at 250/s (approximating the rates of the SMSP and SPEAK strategies), and in another implementation the cycles were repeated at 833/s. The cutoff frequency of the lowpass filters in the envelope detectors was 200 Hz for both implementations. This choice matched the cutoff frequency used in the SMSP and SPEAK strategies. The additional processors evaluated with these subjects included the clinical SPEAK processor used by the subjects in their daily lives and several implementations of CIS processors using different rates of stimulation, different numbers of channels, different polarities of stimulus pulses, different update orders, and different ranges of frequencies spanned by the bandpass filters. (The *n*-of-*m* variations also included processors with a 350- to 5,500-Hz or a 350- to 9,500-Hz range for the bandpass filters.) The tests included identification of consonants in an /a/-consonant-/a/ context for recorded male and female speakers.

In broad terms, the results indicated significantly better performance with the higher-rate *n*-of-*m* processor using the extended frequency range than with the clinical SPEAK processor. This better performance was obtained with no more than a few hours of experience with the *n*-of-*m* processor compared with more than a year of daily use of the SPEAK processor by each of the subjects. The *n*-of-*m* processor also produced the best performance among the tested processors for one of the subjects and the female speaker. Scores across subjects and speakers with the high-rate, extended-frequency range *n*-of-*m* processor were not significantly different from the scores for a six-channel CIS processor also using the extended frequency range and a stimulus rate of 833 pulses/s/electrode. Scores for these two processors were significantly higher than the SPEAK processor or a six-channel CIS processor using the narrower frequency range for the bandpass filters. Scores for these latter two processors were not significantly different.

These comparisons indicate the value of relatively high rates of stimulation and of an extended frequency range for the bandpass filters. They also indicate that *n*-of-*m* processors can be competitive with or better than CIS or SPEAK processors, particularly for certain subjects and speakers.

The *n*-of-*m* approach may allow the representation of spectral detail through use of *m* sites of stimulation, without exceeding the transmission rate limits of transcutaneous links. For example, a 6-of-18 *n*-of-*m* processor may convey more spectral information than a 6-of-6 CIS processor if more than six electrodes (or six bipolar pairs of electrodes) are perceptually separable. In addition, for the same *m* and stimulus rate, an *n*-of-*m* processor is likely to have lower electrode interactions than an *m*-of-*m* (CIS) processor, especially for large values of *m* or close spacing of the intracochlear electrodes. So long as *n* is not too low, the *n*-of-*m* approach may be the best among the available options for some patients, possibly including patients with limited transcutaneous links or relatively high temporal interactions among electrodes.

The ACE implementation of the *n*-of-*m* strategy has been described as combining the best aspects of the SPEAK strategy with the best aspects of the CIS strategy. In particular, CIS has been described as a strategy that emphasizes the representation of temporal variations in speech through relatively rapid rates of stimulation, whereas SPEAK has been described as a strategy that emphasizes the representation of channel or spectral cues through use of a relatively large number of electrodes. The combination of an *n*-of-*m* approach, as in SPEAK, with relatively high rates of stimulation, as typically used in CIS processors, may support especially high levels of speech reception.

Although relatively rapid rates of stimulation can produce improvements in the performance of *n*-of-*m* processors, as demonstrated in the study of Lawson *et al.* and elsewhere (studies cited in Patrick, 1997), the characterization of CIS as using high rates and relatively few electrodes is at least somewhat simplistic. CIS processors may use all of the electrodes available for a particular implant. The first CIS processors were implemented for use with the Ineraid implant and its six electrodes. However, subsequent implementations have addressed 21 of the 22 available electrodes in the Nucleus implant (in the study of Lawson *et al.*) and all 12 of the available electrodes in the Med El COMBI 40+ implant. CIS also may combine rapid rates of stimulation with a large number of electrodes. However, such combinations have been no more effective than combinations of rapid rates with four to six electrodes, as described earlier. This may change with new electrode designs, which may increase the number of perceptually distinct sites of stimulation compared with the number available with the current designs. CIS processors coupled with selective electrodes may well support significant gains in performance with the addition of channels and electrodes beyond six.

The options among SPEAK, high-rate *n*-of-*m* (ACE), and CIS in the new Cochlear Ltd. CI24M device, and the options between high-rate *n*-of-*m* and CIS in the Med El COMBI 40 and COMBI 40+ devices, may help patients to achieve a better outcome than would be possible with one option only. Results from various studies have established the value of the *n*-of-*m* approach and have demonstrated that it can be superior to SPEAK and CIS for some patients and talkers, at least in quiet conditions. Further improvements in the performance of *n*-of-*m* processors may be produced with different rules for selecting the envelope signals to be represented in each cycle of stimulation or with application of the strategy in conjunction with electrodes that have a greater spatial selectivity of stimulation than current electrodes.

Simultaneous Analog Stimulation Strategy

Advanced Bionics has developed two variations of CA processors for use with its Clarion implant. Both variations use back-end compression, in contrast to the front-end compression of the prior CA implementations in the Ineraid and UCSF/Storz devices. The AGC that feeds the bank of bandpass filters in the Clarion implementations has relatively long attack and release times and a relatively low compression ratio. As in CIS processors that use an AGC, this greatly reduces the spectral distortions produced by the AGC stage in the prior CA implementations. As in the CIS and other strategies, bandpass channel outputs in the CA implementations of the Clarion device are individually mapped onto stimulus amplitudes using a nonlinear compression function. Such channel-by-channel mapping may help to produce normal or nearly normal growths of loudness within channels and use most or all of the dynamic range of perception available at each electrode position.

In the first CA implementation used with the Clarion, channel outputs were directed to offset radial pairs of bipolar electrodes. As many as eight channels and associated pairs of electrodes could be used. The offset radial electrodes were modeled after the original UCSF design (Loeb *et al.*, 1983), which could support high spatial specificity of stimulation for cochleas with good survival of neural processes peripheral to the ganglion cells and with positioning of the bipolar pairs immediately adjacent to the peripheral processes.

Such selectivity of stimulation could reduce electrode interactions compared with those produced with monopolar stimulation. Unfortunately, survival of processes peripheral to the ganglion cells is rare in the deaf human cochlea (Hinojosa and Marion, 1983) (see Chapter 6). Also, results of recent studies have demonstrated that a mechanical memory for the curvature of the first and second turns in the average cochlea, as implemented in the original UCSF design, is not sufficient to ensure close apposition of the electrodes to the osseous spiral lamina (where surviving peripheral processes would reside) or to the inner wall of the scala tympani. New approaches for placement of electrodes next to the inner wall of the scala tympani are described in Chapter 6.

Although the first CA implementation used with the Clarion addressed the likely problems associated with front-end compression, it probably did not address the likely problems associated with the interactions that can be produced with simultaneous stimulation across electrode positions. Conditions for selective stimulation were not met with the offset radial electrodes. Quite high levels of stimulation often were required to produce auditory percepts with these electrodes, which probably reflected a focused electrical field coupled with a relatively long distance to excitable neurons (in the spiral ganglion or modiolus), along with large shunting currents between the closely spaced electrodes of each bipolar pair. Many patients could not be stimulated at all within the current and charge limits of the device using the offset radial electrodes.

The fact that many patients could not be stimulated with the offset radial electrodes led Advanced Bionics to reassign the outputs of its implanted receiver/stimulator to address monopolar electrodes or "enhanced bipolar" electrodes, which had a greater spacing between the electrodes in each bipolar pair (about 1.7 mm). Seven such enhanced bipolar electrodes could be addressed.

The Clarion device implements a CIS strategy that addresses monopolar electrodes and a variation of a CA strategy that addresses the enhanced bipolar electrodes. To distinguish this variation of a CA strategy from the prior variation (which addressed up to eight pairs of offset radial electrodes), Advanced Bionics has called it the SAS strategy. Like the CIS strategy, the SAS strategy addresses possible weaknesses of the original CA strategy as implemented in the Ineraid and UCSF/Storz devices. In contrast to the CA strategy of the Ineraid device, the SAS strategy uses the enhanced bipolar electrodes, which may be more selective in stimulating neurons along the longitudinal dimension of the cochlea compared with monopolar electrodes. (The enhanced bipolar electrodes also may be more selective than the offset radial electrodes of the UCSF/Storz device, at least for typical neural survival patterns and placements of electrodes.) The use of back-end compression in the SAS strategy addresses the likely problems associated with front-end compression.

The SAS strategy also uses more channels and electrodes than the prior CA processors of the Ineraid and UCSF/Storz devices. In the default condition, SAS outputs are directed to seven pairs of enhanced bipolar electrodes, whereas the CA outputs were directed to four monopolar electrodes in the Ineraid device or to four offset-radial bipolar electrodes in the UCSF/Storz device. If the spatial specificity of enhanced bipolar electrodes is relatively high, the additional channels in SAS may help.

Evaluation of the SAS strategy is just beginning. In preliminary studies, attempts were made to fit SAS in groups of newly implanted patients. The results showed that the dynamic ranges of percepts from threshold to loud sounds could be spanned for most of the patients using analog stimuli and the enhanced bipolar electrodes. Patients who could be fit with SAS and CIS strategies were asked to indicate a preference between the two. Some indicated a preference for SAS, whereas others indicated a preference for CIS. The proportion of patients indicating a preference for SAS was about 50% among a group implanted in Hannover, Germany (Battmer *et al.*, 1997b), and about 30% among a group implanted at various medical centers in the United States (Osberger, 1998; Kessler, 1998).

The patients in these studies were allowed to use their preferred strategy after the initial fittings and then tested periodically with both strategies. The results suggest a relationship between initial preference and performance, but this impression needs to be verified in a crossover study that provides controls for possible order and experience effects. Such a study is in progress, and results from fully controlled comparisons of a SAS strategy using enhanced bipolar electrodes versus a CIS strategy using monopolar electrodes should be available in the near future.

Convergence of Findings

The developments outlined in this chapter are good news for recipients of cochlear implants. Applications of the CIS and SPEAK processing strategies have produced large improvements in speech reception performance compared with prior strategies. Realistic expectations for prospective patients can be higher than was the case before these strategies became available for clinical use.

Progressive improvements in performance

were obtained in the series of strategies developed for the Nucleus implant as more information was added to the representation for the feature-extraction strategies and when the feature-extraction approach was abandoned in favor of a filterbank approach. The strategies in this series included the F0/F2, F0/F1/F2, MPEAK, and SMSP/SPEAK strategies, each with better performance than its predecessor.

In another line of developments, a filterbank approach also was superior to a strategy that was modeled after the analysis portion of a channel vocoder. In particular, the filterbank approach of the CIS strategy was superior to variation 1 of the IP strategy, in which the features of voiced or unvoiced boundaries and the fundamental frequency of voiced speech sounds were extracted and represented.

Assumptions about how speech is produced and perceived, as used in the design of vocoder systems, are not necessary for the design of effective processing strategies for cochlear implants. Strategies based on such assumptions have produced relatively low levels of performance, especially for listening to speech in competition with noise.

The initial comparisons between the CIS and CA strategies demonstrated large improvements in speech reception scores with the former. This better performance may have been produced by one or more of the following: a reduction in electrode interactions through the use of nonsimultaneous stimuli, the use of five or six channels instead of four, representation of rapid envelope variations through relatively high cutoff frequencies for the lowpass filters in the envelope detectors and rates of stimulation at least twice as high as the cutoff frequencies, or preservation of amplitude cues with channel-by-channel compression and logarithmic or power-law mapping functions.

The new analog strategy of SAS also addresses possible weaknesses of the prior CA strategies. It addresses likely problems associated with front-end compression by using back-end compression on an channel-by-channel basis. To the extent that a high spatial specificity of stimulation is achieved with the enhanced bipolar electrodes, it also may address the likely problem of electrode interactions with the prior strategies. Controlled comparisons of SAS with CIS are underway in a crossover study involving multiple medical centers in the United States. The results may show that SAS is competitive with or better than CIS for some patients.

Various *n*-of-*m* strategies available with the Cochlear Ltd. CI24M implant and the Med El COMBI 40 and COMBI 40+ implants may provide higher speech reception scores for some patients than the SPEAK or CIS strategies. The higher stimulus rates used with the *n*-of-*m* implementation in the CI24M device, compared with the rates used with the SPEAK implementation, may prove to be beneficial. The selection of the highest envelope signals (and the rejection of the lowest envelope signals) may help in listening to speech in noise or for patients with relatively large temporal interactions among electrodes. Studies are in progress to evaluate these possibilities.

The available processing strategies for cochlear implants all use a filterbank or waveform approach. Most of the strategies use nonsimultaneous pulses to reduce deleterious effects of electrode interactions. One of the strategies uses simultaneous stimulation across electrodes but does this in conjunction with bipolar electrodes that may provide a greater spatial specificity of stimulation than other configurations of electrodes. All strategies use back-end, channel-by-channel compression. None of the strategies extracts nor represents specific features of speech.

ADDITIONAL CONSIDERATIONS

Importance of Fitting

Large improvements in the speech reception performance of implant systems can be obtained through informed choices of parameters within a particular processing strategy.

In studies with CIS processors, for example, quite large gains in performance have been produced through choices of pulse rate, pulse duration, electrode update order, the range of frequencies spanned by the bandpass filters, and other parameters (Wilson *et al.*, 1995). Although predetermined values for some parameters may be appropriate for virtually all patients, the values of other parameters should be varied over certain ranges to optimize performance for individuals. An objective of current research in several laboratories is to identify values that can be fixed and values that must be varied to approximate optimal performance within and across patients. The results of such research may inform the development of improved and highly efficient fitting procedures. In the interim, it is important to understand that optimal fittings of speech processors cannot be accomplished without parametric manipulations and that control measures are necessary to gauge the effects of the manipulations. Such parametric studies take time, usually much more time than is typically allocated for the fitting of implant systems in clinical settings.

Importance of Strategy Implementations

The performance of a given strategy also can be affected by the quality of its implementation in speech processor hardware and software. As mentioned before, large increases in speech reception scores were produced in the study of Dowell *et al.* (1991) when the MSP implementation of the F0/F1/F2 strategy was substituted for the WSP III implementation of that strategy. In another two studies, Battmer *et al.* (1997a) and Kessler (1997) found significant improvements in speech reception scores when the Clarion version 1.2 implementation of a CIS strategy was substituted for the version 1.1 implementation of that strategy. Seemingly subtle changes in hardware and the programming of that hardware can produce large changes in performance. This fact complicates comparisons of processing strategies, in that one or both of the strategies under test may not be implemented in the best possible way. For example, an apparent superiority of one of the strategies may be an artifact of a high-quality implementation of that strategy and a less than optimal implementation of the other strategy. This is a particular problem in comparisons involving CIS or CIS-like strategies, because those strategies have been implemented in different ways in a variety of commercial devices (including the Philips LAURA device, the Advanced Bionics Clarion device, the Med-El Combi 40 and Combi 40+ devices, and the Cochlear Ltd. CI24M device) and in a variety of custom processors for laboratory studies.

The details of the implementation are important. Examples of ways in which the implementation can go awry include use of microphones with poor frequency response or high levels of noise, use of amplifier and AGC circuits with low dynamic ranges or high levels of noise, use of digital filters with a reduced number of elements compared with conventional and well-behaved digital filters (a reduced number of elements have been used in devices with small memories or slow digital–signal–processing chips), use of current sources that are especially noisy, use of current sources that saturate or begin to saturate in the dynamic ranges of the electrodes for some or all patients (current sources saturate when the commanded current requires a voltage at the electrodes that approaches or is greater than the voltage limit of the device), and an excessive amount of digital or switching noise at the electrodes. Any one of these can degrade or destroy the performance of an otherwise good strategy.

Importance of the Patient Variable

Data indicating the importance of the patient variable are presented in Fig. 7.18. This figure shows a scatter plot of NU-6 word scores from the within-subject comparisons of CA and CIS processors described previously (scores are from Fig.

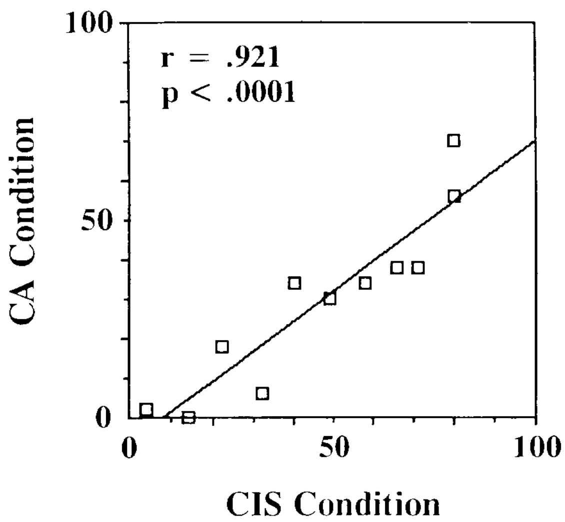

FIG. 7.18. Scatter plot of Northwestern University Auditory Test 6 scores for compressed analog (CA) and continuous interleaved sampling (CIS) processors. Each point represents scores for one subject, and the Pearson correlation coefficient, level of significance, and regression line are shown. (Adapted from Wilson *et al.*, 1993, with permission.)

7.15). These scores were obtained with careful and complete fittings of the two strategies and with high-quality implementations for both. Notice that relatively low scores for one strategy are associated with relatively low scores for the other strategy and vice versa. The data points in Fig. 7.18 are highly correlated (r = 0.92), indicating that 85% of the variance in the data is explained by the subject variable (r^2 = 0.85). Correlations for other tests not obviously distorted by ceiling effects also are quite high for these subjects and processors, in a range between 0.87 and 0.92 (Wilson *et al.*, 1993).

Although outcomes can be improved with a change in processing strategy, the patient variable can have an even greater effect. A patient who enjoys a high ranking among patients with one processor is likely to retain that ranking with another well-implemented and well-fit processor, and a patient who has a low ranking with one processor also is likely to retain that low ranking with another processor. Identification of the factor or factors that underlie the effects of the patient variable may lead to the development of reliable prognostic tests for prospective patients. Knowledge of the factor or factors also may help in the design of better implant systems, that take the factor or factors into account and minimize or eliminate their deleterious effects for patients who otherwise would have relatively poor outcomes.

TRENDS AND CHALLENGES

One of the most striking findings from research on cochlear implants is that the range of performance across patients is large, even with the new processing strategies and current implant systems. Some patients score at or near 100% correct on standard audiologic tests of sentence and word recognition, whereas other patients obtain low scores using an identical speech processor and electrode array. Perhaps the greatest single challenge in improving implants is to identify the mechanisms underlying this variability in outcomes and to use that knowledge in developing new ways to help patients at the low end of the performance spectrum.

Another challenge is to provide more help for all implant users in listening to speech in competition with other speakers or background noise. Speech reception scores even for the best patients are markedly reduced in such situations.

Fortunately, multiple possibilities for improvements in implant systems are under investigation, including new electrode designs for placing electrode contacts close to the inner wall of the scala tympani and thereby reducing thresholds, increasing dynamic range, and increasing the spatial selectivity of stimulation; new ways of representing temporal variations within channels, such as use of very high carrier rates or high-rate conditioner pulses; and coordinated stimulation of bilateral implants, designed to restore sound lateralization abilities and the signal-to-noise advantages that accompany such abilities. These and other possibilities are described in detail in several recent articles (Clark, 1995; Klinke and Hartmann, 1997; Lenarz, 1997; Wilson, 1997 and 1999).

Research is in progress at several centers to identify the mechanisms underlying the wide variation in outcomes with current implant systems. Application of the results may lead to better performance through a reduction in the variation with improved devices. Implants of the near future are likely to be much better than those of today.

REFERENCES

Aubert L, Nicolai J, Staller S, Shaw S. Results of a premarket evaluation of advanced speech encoders in the Nucleus 24 cochlear implant system [Abstract 68]. Presented at the *Fourth European Symposium on Paediatric Cochlear Implantation*; s'Hertogenbosch, The Netherlands, 1998.

Battmer RD, Feldmeier I, Kohlenberg A, Lenarz T. Performance of the new Clarion speech processor 1.2 in quiet and in noise. *Am J Otol* 1997a;18:S144–S146.

Battmer RD, Haake P, Lenarz T. Comparison study of the continuous interleaved sampling (CIS) and compressed analog (CA) strategies, using the CLARION cochlear implant system [Abstract 120]. Presented at the *Vth International Cochlear Implant Conference*; New York, NY, 1997b.

Blamey PJ, Dowell RC, Clark GM, Seligman PM. Acoustic parameters measured by a formant-estimating speech processor for a multiple-channel cochlear implant. *J Acoust Soc Am* 1987;82:38–47.

Böex C, Pelizzone M, Montandon P. Improvements in speech recognition with the CIS strategy for the Ineraid multichannel intracochlear implant. In: Hochmair IJ, Hochmair ES, eds. *Advances in cochlear implants*. Vienna: Manz, 1994:136–140.

Böex C, Pelizzone M, Montandon P. Speech recognition with a CIS strategy for the Ineraid multichannel cochlear implant. *Am J Otol* 1996;17:61–68.

Busby PA, Tong YC, Clark GM. The perception of temporal modulations by cochlear implant patients. *J Acoust Soc Am* 1993;94:124–131.

Clark GM. The University of Melbourne-Nucleus multi-electrode cochlear implant. *Adv Otorhinolaryngol* 1987;38:1–189.

Clark GM. Cochlear implants: future research directions. *Ann Otol Rhinol Laryngol* 1995;104[Suppl 166]:22–27.

Clark GM, Tong YC, Patrick JF. *Cochlear prostheses*. Edinburgh: Churchill Livingstone, 1990.

Cohen NL, Waltzman SB, Fisher SG, *et al.* A prospective, randomized study of cochlear implants. *N Engl J Med* 1993;328:233–237.

Crosby PA, Daly CN, Money DK, Patrick JF, Seligman PM, Kuzma JA. Cochlear implant system for an auditory prosthesis. US patent 4532930, 1985.

Dorman MF. Speech perception by adults. In: Tyler RS, ed. *Cochlear implants: audiological foundations*. San Diego: Singular Publishing Group, 1993:145–190.

Dorman MF, Hannley MT, Dankowski K, Smith L, McCandless G. Word recognition by 50 patients fitted with the Symbion multichannel cochlear implant. *Ear Hear* 1989;10:44–49.

Dorman MF, Loizou PC. Changes in speech intelligibility as a function of time and signal processing strategy for an Ineraid patient fitted with continuous interleaved sampling (CIS) processors. *Ear Hear* 1997a; 18:147–155.

Dorman MF, Loizou PC. Mechanisms of vowel recognition for Ineraid patients fit with continuous interleaved sampling processors. *J Acoust Soc Am* 1997b;102:581–587.

Dorman MF, Smith L, Parkin JL. Loudness balance between acoustic and electric stimulation by a patient with a multichannel cochlear implant. *Ear Hear* 1993;14:290–292.

Dowell RC, Dawson PW, Dettman SJ, *et al.* Multichannel cochlear implantation in children: a summary of current work at the University of Melbourne. *Am J Otol* 1991;12[Suppl 1]:137–143.

Dowell RC, Mecklenberg DJ, Clark GM. Speech recognition for 40 patients receiving multichannel cochlear implants. *Arch Otolaryngol Head Neck Surg* 1986; 112:1054–1059.

Dowell RC, Seligman PM, Blamey PJ, Clark GM. Speech perception using a two-formant 22-electrode cochlear prosthesis in quiet and in noise. *Acta Otolaryngol* 1987a;104:439–446.

Dowell RC, Seligman PM, Blamey PJ, Clark GM. Evaluation of a two-formant speech-processing strategy for a multichannel cochlear prosthesis. *Ann Otol Rhinol Laryngol* 1987b;96[Suppl 128]:132–134.

Eddington DK. Speech discrimination in deaf subjects with cochlear implants. *J Acoust Soc Am* 1980; 68:885–891.

Eddington DK, Dobelle WH, Brackman DE, Mladejovsky MG, Parkin JL. Auditory prosthesis research with multiple channel intracochlear stimulation in man. *Ann Otol Rhinol Laryngol* 1978;87[Suppl 53]:1–39.

Favre E, Pelizzone M. Channel interactions in patients using the Ineraid multichannel cochlear implant. *Hear Res* 1993;58:79–90.

Fishman K, Shannon RV, Slattery WH. Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. *J Speech Hear Res* 1997;40:1201–1215.

Flanagan JL. *Speech analysis, synthesis and perception*, 2nd ed. Berlin: Springer-Verlag, 1972.

Gantz BJ. Cochlear implants: an overview. *Adv Otolaryngol Head Neck Surg* 1987;1:171–200.

Gantz BJ, McCabe BF, Tyler RS, Preece JP. Evaluation of four cochlear implant designs. *Ann Otol Rhinol Laryngol* 1987;96[Suppl 128]:145–147.

Gantz BJ, Tyler RS, Knutson JF, *et al.* Evaluation of five different cochlear implant designs: audiologic assessment and predictors of performance. *Laryngoscope* 1988;98:1100–1106.

Hess W. *Pitch determination of speech signals.* Berlin: Springer-Verlag, 1983.

Hinojosa R, Marion M. Histopathology of profound sensorineural deafness. *Ann NY Acad Sci* 1983; 405:459–484.

Kessler DK. Clinical investigation of the CLARION Multi-Strategy—adult patient performance with Clarion version 1.2 [Abstract 118]. Presented at the *Vth International Cochlear Implant Conference*; New York, NY, 1997.

Kessler DK. New directions in speech processing II: The electrode connection [Abstract 159]. Presented at the *Fourth European Symposium on Paediatric Cochlear Implantation*; s'Hertogenbosch, The Netherlands, 1998.

Kiefer J, Müller J, Pfennigdorff T, *et al.* Speech understanding in quiet and in noise with the CIS speech coding strategy (Med-El Combi-40) compared to the multipeak and spectral peak strategies (Nucleus). *ORL J Otorhinolaryngol Relat Spec* 1996;58:127–135.

Klinke R, Hartmann R. Basic neurophysiology of cochlear implants. *Am J Otol* 1997;18:S7–S10.

Lawson DT, Wilson BS, Finley CC. New processing strategies for multichannel cochlear prostheses. *Prog Brain Res* 1993;97:313–321.

Lawson DT, Wilson BS, Zerbi M, Finley CC. Speech processors for auditory prostheses. First Quarterly Progress Report, NIH project N01-DC-5-2103. Bethesda: Neural Prosthesis Program, National Institutes of Health, 1995.

Lawson DT, Wilson BS, Zerbi M, Finley CC. Speech processors for auditory prostheses. Third Quarterly Progress Report, NIH project N01-DC-5-2103. Bethesda: Neural Prosthesis Program, National Institutes of Health, 1996.

Lenarz T. Cochlear implants—What can be achieved? *Am J Otol* 1997;18:S2–S3.

Loeb GE, Byers CL, Rebscher SJ, *et al.* Design and fabrication of an experimental cochlear prosthesis. *Med Biol Eng Comput* 1983;21:241–254.

Loizou PC. Mimicking the human ear: an overview of signal processing strategies for cochlear prostheses. *IEEE Sig Proc Mag* 1998;15:101–130.

Loizou P, Graham S, Dickens J, Dorman M, Poroy O. Comparing the performance of the SPEAK strategy (Spectra 22) and the CIS strategy (Med-El) in quiet and in noise [Poster abstract 64]. Presented at the *1997 Conference on Implantable Auditory Prostheses*; Pacific Grove, CA, 1997.

McDermott HJ, McKay CM, Vandali AE. A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant. *J Acoust Soc Am* 1992;91:3367–3391.

McDermott HJ, Vandali AE. Spectral maxima sound processor. US patent 5597380, 1997.

McKay CM, McDermott HJ, Clark GM. Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees. *J Acoust Soc Am* 1994;96:2664–2673.

McKay C, McDermott H, Vandali A, Clark G. Preliminary results with a six spectral maxima sound processor for the University of Melbourne/Nucleus multiple-electrode cochlear implant. *J Otolaryngol Soc Aust* 1991;6:354–359.

McKay CM, McDermott HJ, Vandali AE, Clark GM. A comparison of speech perception of cochlear implantees using the Spectral Maxima sound Processor (SMSP) and the MSP (MULTIPEAK) processor. *Acta Otolaryngol* 1992;112:752–761.

Merzenich MM. UCSF cochlear implant device. In: Schindler RA, Merzenich MM, eds. *Cochlear implants.* New York: Raven Press, 1985:121–130.

Merzenich MM, Rebscher SJ, Loeb GE, Byers CL, Schindler RA. The UCSF cochlear implant project. *Adv Audiol* 1984;2:119–144.

Millar JB, Blamey PJ, Tong YC, Patrick JF, Clark GM. Speech perception. In: Clark GM, Tong YC, Patrick JF, eds. *Cochlear prostheses.* Edinburgh: Churchill Livingstone, 1990:41–67.

Millar JB, Tong YC, Clark GM. Speech processing for cochlear implant prostheses. *J Speech Hear Res* 1984;27:280–296.

Moore BCJ. Speech coding for cochlear implants. In: Gray RF, ed. *Cochlear implants.* San Diego: College-Hill Press, 1985:163–179.

Osberger MJ. New directions in speech processing. I: Patient performance with simultaneous analog stimulation (SAS) [Abstract 158]. Presented at the *Fourth European Symposium on Paediatric Cochlear Implantation*; s'Hertogenbosch, The Netherlands, 1998.

O'Shaughnessy D. *Speech communication: human and machine.* Reading, MA: Addison-Wesley, 1987.

Papamichalis PE. *Practical approaches to speech coding.* Englewood Cliffs, NJ: Prentice-Hall, 1987.

Parkins CW. Cochlear prostheses. In: Altschuler RA, Hoffman DW, Bobbin RP, eds. *Neurobiology of hearing: the cochlea.* New York: Raven Press, 1986:455–473.

Patrick JF. The evolution of speech coding strategies [Abstract 5]. Presented at the *Vth International Cochlear Implant Conference*; New York, NY, 1997.

Patrick JF, Clark GM. The Nucleus 22-channel cochlear implant system. *Ear Hear* 1991;12[Suppl 1]:3S–9S.

Patrick JF, Seligman PM, Clark GM. Engineering. In: Clark GM, Cowan RSC, Dowell RC, eds. *Cochlear implantation for infants and children: advances.* San Diego: Singular Publishing Group, 1997:125–145.

Patrick JF, Seligman PM, Money DK, Kuzma JA. Engineering. In: Clark GM, Tong YC, Patrick JF, eds. *Cochlear prostheses.* Edinburgh: Churchill Livingstone, 1990:99–124.

Pelizzone M, Böex-Spano C, Sigrist A, *et al.* First field trials with a portable CIS processor for the Ineraid multichannel cochlear implant. *Acta Otolaryngol* 1995;115:622–628.

Pfingst BE. Stimulation and encoding strategies for cochlear prostheses. *Otolaryngol Clin North Am* 1986;19:219–235.

Rabiner LR, Shafer RW. *Digital processing of speech signals.* Englewood Cliffs, NJ: Prentice-Hall, 1978: 129–130.

Shannon RV, Adams DD, Ferrel RL, Palumbo RL, Grandgenett M. A computer interface for psychophysical and speech research with the Nucleus cochlear implant. *J Acoust Soc Am* 1990;87:905–907.

Skinner MW, Clark GM, Whitford LA, *et al.* Evaluation of a new spectral peak (SPEAK) coding strategy for the Nucleus 22 channel cochlear implant system. *Am J Otol* 1994;15[Suppl 2]:15–27.

Skinner MW, Holden LK, Holden TA, *et al.* Performance of postlinguistically deaf adults with the Wearable Speech Processor (WSP III) and Mini Speech Processor (MSP) of the Nucleus multi-electrode cochlear implant. *Ear Hear* 1991;12:3–22.

Tye-Murray N, Lowder M, Tyler RS. Comparison of the F0F2 and F0F1F2 processing strategies for the Cochlear Corporation cochlear implant. *Ear Hear* 1990;11:195–200.

Tye-Murray N, Tyler RS. Auditory consonant and word recognition skills of cochlear implant users. *Ear Hear* 1989;10:292–298.

Tyler RS, Lowder MW, Parkinson AJ, Woodworth GG, Gantz BJ. Performance of adult Ineraid and Nucleus cochlear implant patients after 3.5 years of use. *Audiology* 1995;34:135–144.

Tyler RS, Moore BCJ. Consonant recognition by some of the better cochlear-implant patients. *J Acoust Soc Am* 1992;92:3068–3077.

Tyler RS, Preece JP, Lansing CR, Otto SR, Gantz BJ. Previous experience as a confounding factor in comparing cochlear-implant processing schemes. *J Speech Hear Res* 1986;29:282–287.

Tyler RS, Tye-Murray N. Cochlear implant signal-processing strategies and patient perception of speech and environmental sounds. In: Cooper H, ed. *Cochlear implants: a practical guide*. San Diego: Singular Publishing Group, 1991:58–83.

White MW. Compression systems for hearing aids and cochlear prostheses. *J Rehab Res Dev* 1986;23:25–39.

White MW, Merzenich MM, Gardi JN. Multichannel cochlear implants: channel interactions and processor design. *Arch Otolaryngol* 1984;110:493–501.

Wilson BS. Signal processing. In: Tyler RS, ed. *Cochlear implants: audiological foundations*. San Diego: Singular Publishing Group, 1993:35–85.

Wilson BS. The future of cochlear implants. *Br J Audiol* 1997;31:205–225.

Wilson BS. New directions in implant design. In: Waltzman SB, Cohen N, eds. *Cochlear implants*. New York: Thieme Medical and Scientific Publishers, 1999:43–56.

Wilson BS, Finley CC, Farmer JC Jr, *et al.* Comparative studies of speech processing strategies for cochlear implants. *Laryngoscope* 1988a;98:1069–1077.

Wilson BS, Finley CC, Lawson DT. Speech processors for auditory prostheses. Seventh Quarterly Progress Report, NIH project N01-NS-3-2356. Bethesda: Neural Prosthesis Program, National Institutes of Health, 1985.

Wilson BS, Finley CC, Lawson DT, Wolford RD. Speech processors for cochlear prostheses. *Proc IEEE* 1988b;76:1143–1154.

Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM. Better speech recognition with cochlear implants. *Nature* 1991a;352:236–238.

Wilson BS, Lawson DT, Finley CC, Wolford RD. Coding strategies for multichannel cochlear prostheses. *Am J Otol* 1991b;12[Suppl 1]:56–61.

Wilson BS, Lawson DT, Finley CC, Wolford RD. Importance of patient and processor variables in determining outcomes with cochlear implants. *J Speech Hear Res* 1993;36:373–379.

Wilson BS, Lawson DT, Zerbi M. Advances in coding strategies for cochlear implants. *Adv Otolaryngol Head Neck Surg* 1995;9:105–129.

Zeng F-G, Shannon RV. Loudness balance between acoustic and electric stimulation. *Hear Res* 1992;60:231–235.

Cochlear Implants: Principles & Practices, edited by John K. Niparko, Karen Iler Kirk, Nancy K. Mellon, Amy McConkey Robbins, Debara L. Tucci, and Blake S. Wilson. Lippincott Williams & Wilkins, Philadelphia © 2000.

III. Plans for the next quarter

Our plans for the next quarter include the following:

- Ongoing studies with subject SR2. We expect that studies for the next quarter will include completion of work in progress to evaluate (1) use of the TIMIT speech database as a source of difficult sentences for sensitive measures of speech reception by a high-performance subject and (2) "conditioner pulses" processors. If use of the TIMIT database proves to be useful (e.g., if scores show low list-to-list variations), then we also expect to apply the database in further studies of rate/lowpass filter effects and in further studies of effects of manipulations in mapping functions.
- Continued development of a new strategy, designed to mimic closely the nonlinear processing in the peripheral auditory system, including the strong and nearly instantaneous compression at the basilar membrane for sound pressure levels above 35-40 dB and the strong and noninstantaneous (with multiple time constants) compression that occurs at the synapse between inner hair cells and type I fibers of the auditory nerve. Development of the new strategy includes extensive modeling in MATLAB, and a MATLAB system has been purchased in part for support of this effort.
- Continued preparation for studies with recipients of CI24M implants on both sides, and with recipients of COMBI 40+ implants on both sides.
- A visit by Marian Zerbi to assist in the above preparation.
- Resumption of studies with recipients of CI24M implants on both sides, using new software and hardware that will allow evaluation of strategies that provide coordinated stimuli to the two sides, to preserve a representation of localization cues, and measures of sensitivities to interaural timing and amplitude differences using adaptive procedures.
- A visit by Jan Kiefer and Thomas Pfennigdorff, of the J.W. Goethe Universität in Frankfurt, on February 1 and 2, for further discussions on combined electric and acoustic stimulation of the same cochlea and for further development of plans for cooperative studies between the university and RTI. (This visit follows one by Blake Wilson to Frankfurt; see Introduction.)
- Presentation of project results in an invited lecture at the *6th International Cochlear Implant Conference*, Miami Beach, FL February 3-5, 2000.
- Visits by Joachim Müller of the Julius-Maximilians Universität in Würzburg, Peter Nopp of the Med El company in Innsbruck, and Arturs Lorens of the Institute of Physiology and Pathology of Hearing in Warsaw, for between two days (Müller) and one week (Nopp and Lorens) following the Miami conference.
- A *Mini Symposium on Cochlear Implants* at RTI, held in conjunction with, and in honor of, the above visitors (February 7).
- Studies with Ineraid subject SR15 during the week beginning on February 7. We expect that the studies will include (a) longitudinal measures with her portable CIS (CIS-Link) processor, (b) measures of consonant identification for CIS processors using a wide range of compression functions, and (c) evaluation of combinations of low stimulus rates and low cutoff frequencies for the lowpass filters in the envelope detectors in CIS processors, as suggested by the results from prior psychophysical scaling experiments with this subject. (The prior scaling results indicated abnormally low asymptotes in pitch judgements with increases in rate of stimulation or with increases in modulation frequency for SAM pulse trains, and the prior results also indicated an abnormally low sensitivity to modulation depth for SAM pulse trains.)
- Completion of the Access database mentioned in the Introduction.
- Continued analysis of psychophysical, speech reception, and evoked potential data from current and prior studies.
- Continued preparation of manuscripts for publication.

IV. Announcement

Stefan Brill began in this quarter the postdoctoral appointment first mentioned in Quarterly Progress Report 3 for this project. The anticipated term of the appointment is two years.

As noted in QPR 3, Stefan's Ph.D. work, at the University of Vienna and conducted under the guidance of and in cooperation with Erwin Hochmair at the University of Innsbruck, was in the design and evaluation of speech processing strategies for cochlear prostheses. Stefan's experience includes work with implant patients, digital signal processing, assembly-language programming of DSP chips, a wide range of additional programming languages, and teaching. He has presented papers at many international conferences on cochlear implants and related topics. He also recently published a paper in the American Journal of Otology, on "Optimization of channel number and stimulation rate for the fast continuous interleaved sampling strategy in the COMBI 40+." He is a member of the International Functional Electrical Stimulation Society.

Stefan's main work at RTI will involve studies with recipients of bilateral implants. He will play a major role in upcoming studies with recipients of COMBI 40+ implants on both sides, in cooperation with the University of Würzburg, and with recipients of CI24M implants on both sides, in cooperation with the University of Iowa.

We are very pleased to have Stefan as a member of the RTI team.

V. Acknowledgments

We thank subjects SR2 and SR10 for their participation in the studies of this quarter.

Appendix 1. Summary of reporting activity for this quarter

Reporting activity for this quarter, covering the period of October 1 through December 31, 1999, included the following:

Presentations

Wilson BS: Speech processors for auditory prostheses. Invited lecture presented at the *30th Neural Prosthesis Workshop*, Bethesda, MD, October 12-14, 1999.

Wilson BS: Psychophysical measures and speech understanding in bilaterally implanted patients. Invited lecture presented at the *Bilateral Research Meeting* (sponsored by the Med El company), Frankfurt, Germany, December 3, 1999.

Publications

Wilson BS: New directions in implant design. In *Cochlear Implants*, edited by SB Waltzman and N Cohen, Thieme Medical and Scientific Publishers, New York, NY, 2000, pp. 43-56.

Wilson BS: Cochlear implant technology. In *Cochlear Implants: Principles & Practices*, edited by JK Niparko, KI Kirk, NK Mellon, AM Robbins, DL Tucci and BS Wilson, Lippincott Williams & Wilkins, Philadelphia, PA, 2000, pp. 109-119.

Wilson BS: Strategies for representing speech information with cochlear implants. In *Cochlear Implants: Principles & Practices*, edited by JK Niparko, KI Kirk, NK Mellon, AM Robbins, DL Tucci and BS Wilson, Lippincott Williams & Wilkins, Philadelphia, PA, 2000, pp. 129-170.

Niparko JK, Wilson BS: History of cochlear implants. In *Cochlear Implants: Principles & Practices*, edited by JK Niparko, KI Kirk, NK Mellon, AM Robbins, DL Tucci and BS Wilson, Lippincott Williams & Wilkins, Philadelphia, PA, 2000, pp. 103-107.

Niparko JK, Kirk KI, Mellon NK, Robbins AM, Tucci DL, Wilson BS (Eds.), *Cochlear Implants: Principles & Practices*, Lippincott Williams & Wilkins, Philadelphia, PA, 2000.